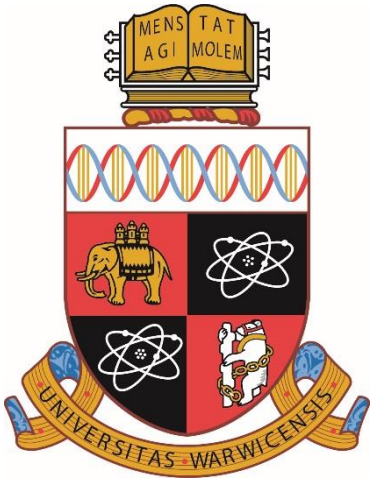


The Development and Optimization of Potential Germanium on Silicon Single Photon Avalanche Diodes

by

Philip Simon Allred

Submitted to the University of Warwick

in partial fulfilment of the requirements

for admission to the degree of

Doctor of Philosophy

Department of Physics

THE UNIVERSITY OF WARWICK

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. All experimental work presented was carried out by the author, or (where stated) by specialists under the author's direction.

Name: ……

Date: ……

Acknowledgments

The work in this thesis would not have been completed without the help, guidance and patience of my supervisors, Prof. David Leadley and Dr. Maksym Myronov. I am very grateful to Dr Maksym Myronov for the proposal, design and epitaxy of Si and Ge based materials by RP-CVD. Also for providing encouragement and plenty of samples to keep me busy, as well as sharing an appreciation for mixed grills.

I would like to thank all the members of the nano Silicon group throughout my time. I owe both Vishal Shah and Stephen Rhead enormous thanks for sharing some of their wisdom. I am grateful for the current members; Olly for pushing my lungs to new limits but being inferior in the weight room, Jamie for the perfect geoguesser distractions (1.6m), Turbo for his chemistry knowledge and vehicular assistance, Dave for the extra curriculars, Gerard for help during the membrane process, and vineet. I would like to thank previous members James, John for help with the TEM, and Cat for the frequent cakes.

I must thank Ind massively for putting up with me for so long. He was able to provide some much needed distractions when they were necessary. I am also very grateful for the support of Stephen Ball throughout, but particularly in the latter stages with some extra gym sessions.

I would also like to thank someone who gave me an immeasurable amount of support through part of my PhD. They may not have realised the significance of her help. But they were there to give me focus through some of the stressful periods, as well as give me some unforgettable experiences. They believed in me at times when perhaps I didn't. I can't thank them enough for the part they played, I was very lucky. I would also like to include their family in these acknowledgements too, as they were very kind to me also. For that I am truly grateful.

Special thanks go to Ally Caldecote, who was there to pick me up during what has proved to be the toughest period of my life. Without the generous supplies of chicken and salmon, and the patience to listen to me, there wouldn't be a thesis to read. She has been there for me when I have really needed.

I would like to thank my friends from home including Pierre Taylor, Rob Tambini, Alex Pope, Adam Lloyd, Edward Vernon, Milky, Sam Rest, and Paul Ramsay. They kept me entertained throughout. Huge thanks go to my family, who have had to see me at my worst. A lot of credit goes to them for getting me to this stage in the first place. While things have been far from smooth at points I don't know where I would be without the help and support I have had.

I would like to thank the EPSRC for funding this study.

Abstract

The work presented in this thesis explores a potential single photon detection technology using Silicon and Germanium, and a possible direction for the future. Instead of the more commonly used III-V materials, the desirable characteristics of each of the Group IV materials is implemented in designing a separate absorption and multiplication region device. Key structural features of the device are investigated and optimised, so that single photon detection in the near infrared is made possible. Growth of these layers is performed using an RP-CVD system, the ultimate industry tool in this field of research. Doping profiles and smooth crystalline growth is implemented using a range of techniques, to produce suitable epitaxial structures which are ideal for further fabrication. Several techniques are used to ensure that the quality of these layers are fully optimised. This optimisation work has resulted in the first single photon detection at a wavelength of 1550 nm, and has also brought the Silicon and Germanium device onto a comparable level to their III-V counterparts at 1330 nm. The superior repetition rate of these Group IV devices also holds an advantage over those designed using InGaAs/InP. The boron doping of Silicon has also been investigated. It has been shown that fully crystalline Silicon boron layers can be produced with boron concentrations ($4.5x10^{20}$ cm^{-3}) that are higher than their solubility limit at 700^{o}C. The reproducibility of these layers, along with quick turnaround, offers an excellent possibility for industrial use, with a significant advantage over other competing growth techniques. In relation to the work on the single photon detection devices, these B-doped layers offer an interesting etch resistant capability. Suspended structures (wires and membranes) have been produced and characterized using synchrotron measurements. Layers have shown small levels of strain, similar to structures made using Germanium, but overall exhibit a flat platform for further growth. This has led to the idea that suspending a single photon detector that incorporates a reflective mirror could enhance detection efficiency.

Publications and Presentations

Papers

1. **P. Allred**, S.D. Rhead, M. Myronov, R.E. Warburton, G. Intermite, G.S. Buller, D.R. Leadley,

   Epitaxial growth challenges of thick Ge-on-Si structures used for single photon avalanche diode applications, In Preparation

2. **P. Allred**, S.D. Rhead, M. Myronov, D.R. Leadley,

   Suspended SiB structures grown using RP-CVD for application to Group IV Single Photon Avalanche Diodes, In Prepartaion

3. V.A. Shah, S.D. Rhead, J. Finch, M. Myronov, J.S. Reparaz, R.J. Morris, N.R. Wilson, V. Kachkanov, I.P. Dolbnya, J.E. Halpin, D. Patchett, P. Allred, G. Colston, K.J.S. Sawhney, C.M. Sotomayor Torres, D.R. Leadley
   Electrical properties and strain distribution of Ge suspended structures
   Solid State Electronics **108**, 13, (2015)

4. S.D. Rhead, J.E. Halpin, V.A. Shah, M. Myronov, D.H. Patchett, **P.S. Allred**, V. Kachkanov, I.P. Dolbnya, J.S. Reparaz, N.R. Wilson, C. M. Sotomayor Torres and D.R. Leadley
   Tensile Strain Mapping in Flat Germanium Membranes
   Applied Physics Letters **104**, 172107 (2014)

5. V.A. Shah, S. Rhead, J Halpin, O. Trushkevych, E. Chávez-Ángel, M. Myronov, J.S. Reparaz, R. Edwards, F. Alzina, A. Shchepetov, S. Kachkanov, N. R. Wilson, I. Dolbnya, D.H. Patchett, **P.S. Allred**, M.J. Prest, P.M. Gammon, M. Prunnila, T.E.Whall, E.H.C. Parker, C.M. Sotomayor Torres, D.R. Leadley

   High Quality Single Crystal Ge Nano-Membranes for Opto-Electronic Integrated Circuitry

   Journal of Applied Physics **115**, 144307 (2014) DOI: 10.1063/1.4870807

6. R.E. Warburton, G. Intermite, M. Myronov, **P. Allred**, D.R. Leadley, K. Gallagher, D. J. Paul, N. J. Pilgrim, L.J.M. Lever, Z. Ikonic, R.W. Kelsall, and G.S. Buller

   Ge-on-Si single photon avalanche diode detectors: Design, modeling, fabrication, and characterization at 1310 and 1550 nm

   IEEE Transactions on Electron Devices, **60**, 2282712 (2013)

Conference Presentations

1. Van Huy Nguyen, M. Myronov, **P. Allred**, J. Halpin A. Dobbie and D.R. Leadley

   Developments in Germanium-on-Silicon epitaxy by reduced pressure chemical vapour deposition

   ULIS 2014, Stockholm, Sweden, April 2014

2. S.D. Rhead, J.E. Halpin, V.A. Shah, M. Myronov, D.H. Patchett, **P.S. Allred**, V. Kachkanov, I.P. Dolbnya, N.R. Wilson and D.R. Leadley

   Tensile Strain Mapping in Flat Germanium Membranes

   7th Intl. Silicon-Germanium Technology and Device Meeting (ISTDM) Singapore, June (2014)

3. **P.S. Allred**, M. Myronov, S.D. Rhead, R. Warburton, G. Intermite, G. Buller and D.R. Leadley

Optimization of Epitaxial Growth for Thick Ge-on-Si Structures used for Single Photon Avalanche Diode Applications

7th Intl. Silicon-Germanium Technology and Device Meeting (ISTDM) Singapore, June (2014)

4. M. Myronov, J. Halpin, S. Rhead, Van Huy Nguyen, D. Patchett, **P. Allred**, J. Foronda, V. Shah, C. Morrison, G. Colston, V. Sivadasan, D.R. Leadley

Silicon-based epitaxy capability at the University of Warwick

UK Silicon Photonics Conference, Southampton, Nov. 2013

5. **P. Allred**, M. Myronov R. Warburton, G. Intermite, K. Gallacher, N. Pilgrim, L. Lever, D.J. Paul, Z. Ikonic, R. Kelsall, G. Buller and D.R. Leadley

Development of Ge-on-Si Single-Photon Avalanche Diodes.

UK Silicon Photonics Conference, Southampton, Nov. 2013

6. G. Intermite, R.E. Warburton, M. Myronov, **P. Allred**, D.R. Leadley, K. Gallagher, D.J. Paul, N.J. Pilgrim, L.J.M. Lever, Z. Ikonic, R.W. Kelsall, and G.S. Buller

Design and performance of a prototype mesa-geometry Ge-on-Si single-photon avalanche diode detector at 1310 nm and 1550 nm wavelengths

Group IV Photonics, Seoul, August 2013

7. O.A. Mironov, A.H.A. Hassan, A. Dobbie, R.J.H. Morris, A. Feher, E. Cizmar, S. Gabani, I.B. Berkutov, J.E. Halpin, S.D. Rhead, **P. Allred**, and D.R. Leadley

New RP CVD grown ultra-high performance selectively B-doped pure-Ge 20 nm

QWs on (100)Si as basis material for post-Si CMOS technology

E-MRS Spring Meeting, Strasburg, May 2013

Glossary of terms

AFM – Atomic force microscope

APD – Avalanche photodiode

CCC – Closed cycle cryostat

CCD – Charge coupled device

CRL – Compound refractive lens

CVD – Chemical vapour deposition

DCR – Dark count rate

FTIR – Fourier transform infrared spectrometry

GS-MBE – Gas source molecular beam epitaxy

HF – Hydrofluoric acid

HH – Heavy hole

IR – Infrared

LH – Light hole

MBE – Molecular beam epitaxy

MOSFET – Metal oxide semiconductor field effect transistor

NEP – Noise equivalent power

PMT – Photomultiplier tube

RP-CVD – Reduced pressure chemical vapour deposition

RSM – Reciprocal space map

SAM – Separate absorption and multiplication

SEM – Scanning electron microscope

SIMS – Secondary ion mass spectrometry

SOI – Silicon on insulator

SPAD – Single photon avalanche diode

SPDE – Single photon detection efficiency

SS-MBE – Solid source molecular beam epitaxy

TDD – Threading dislocation density

TEM – Transmission electron microscope

TMAH – Tetramethylammonium hydroxide

UHV-CVD – Ultra high vacuum chemical vapour deposition

VdP – Van der pauw

XRD – X-ray diffraction

μ - mobility

α – absorption coefficient/electron impact ionization coefficient

β - hole impact ionization coefficient

ρ - resistivity

θ – theta angle

ω – omega angle

F – excess noise factor

M – multiplication gain

E_g – Band gap

λ – wavelength

q_x – in plane reciprocal lattice vector

q_z – out of plane reciprocal lattice vector

V_H – Hall voltage

n_s – sheet density

Table of Contents

3. Experimental Techniques

4. Development of Ge on Si Epitaxial SPAD device

List of Figures

Chapter 1

Introduction to Silicon and Germanium

1.1 The Semiconductor industry

Since the creation of the first transistor at Bell labs in 1947 Silicon has been the dominant semiconductor material used in the industry, accounting for 97% [1]. The semiconductor industry has been estimated to reach over $350 billion in total sales by 2016 [2]. Silicon became the preferred option due to its low cost for integrated circuits. Despite some material properties being inferior compared to other semiconductor materials such as Germanium, the availability and cost advantages of Si significantly outweigh these. Added to this, a high thermal conductivity of Silicon also reduces overheating in the circuits.

Probably the most common application for Silicon as a device is in the metal-oxide-semiconductor-field-effect-transistor (MOSFET). The oxide of Silicon (SiO_2) has excellent insulating qualities making it ideal for device fabrication [3]. The scale of these devices is constantly being reduced so that the density of transistors on an integrated circuit can be increased, resulting in smaller systems. The number of transistors used follows Moore's law, stating that the number of transistors on an integrated circuit doubles roughly every two years.

As previously mentioned, if price was disregarded, Silicon would not be the logical choice to base semiconductors around. For some applications, examples which are not limited to Germanium or an alloy of Silicon and Germanium may be more suitable. However in the case of Germanium, the cost and fragility of Ge substrates prevents them from being a viable option for mass production. Instead Germanium can be grown epitaxially onto the much cheaper Silicon substrate. Unfortunately the cost of growing other materials is not necessarily a cheap process, and may require long growth times. Also the heteroepitaxy (growth of a layer which is different from that preceding it) introduces negative side effects which can result in poor quality layers. As epitaxial growth has been improved techniques have been employed to obtain relatively smooth and defect free layers of materials such as Germanium. With better quality layers, new devices and enhancement of existing ones can be realised. Germanium is used in laser devices [4], solar cells (with III-V materials) [5], and infra-red (IR) photodetectors [6]. It is also grown as an alloy for various applications. Silicon is the most common choice, where Silicon Germanium (SiGe) alloys can be used for example as graded regions in a virtual

substrate [7]. Tin has also become a promising material to use in an alloy with Germanium, as it has been predicted that a direct band gap can be obtained at 12% tin content [8].

1.2 Photon-detection

The detection of single photons has many applications in areas such as biophotonics where the number of photons emitted is often very small [10]. The semiconductor band gap plays an important role in the performance of many devices: one of which is the ability of a semiconductor to absorb photons. For applications in the telecommunication region, wavelengths above 1 μm and up to 1.55 μm are of particular interest. The semiconductor material chosen for an application will be dependent on its band gap, and therefore the wavelengths accessible. Table 1 shows a table of semiconductor and III-V materials and their respective band gap.

Table 1: Band gap values for relevant semiconductor materials. Taken from [10][11].

Material	Band Gap (eV)	Lattice Parameter (A)	Absorption Limit (nm)
Silicon	1.11	5.43	1550
Germanium	0.66	5.66	1000
Indium Phosphide	1.27	5.87	960
InGaAs (variable)	0.73	5.65-6.05	1680

If the energy of a photon is sufficiently large, such that its energy is equal to or exceeds the value of the band gap, it may be absorbed. Shorter wavelengths have a larger energy according to:

$$E = \frac{hc}{\lambda}$$ Equation 1.1

From equation 1.1 it can be seen that for a given band gap there will be a maximum wavelength of photon that can be absorbed. Therefore clearly some materials are unsuitable beyond a specific wavelength so cannot be used in a device which is dependent on detection of such a wavelength. For an alloy of Silicon and Germanium, a mere 5% Silicon can lead to a bandgap of nearly 0.2 eV greater than that of Germanium [12]. This can be problematic for any structures which require small band gaps for detection of near infra-red photons.

The absorption of a photon is not the only important factor in photodetection. Once a photon is absorbed it has to be detected. The absorption of a photon may create an electron-hole pair which can then be detected via an external source. The first device to demonstrate this was the photomultiplier tube (PMT), which was designed in 1930 by Russian physicist Kubetsky. The experiment used primary photoelectrons emitted from a photocathode and accelerated them into consecutive emitters. At each emitter further secondary electrons could be emitted and undergo the same process resulting in a huge amplification. Gains of up to ten thousand were recorded [13].

The process of photodetection has evolved since this invention with the help of semiconductors. Using the same concept as the PMT, avalanche photodiodes have been established as a compact and integratable device. However, this market has been dominated by III-V materials due to their more appropriate band gap when operating at the principal telecommunication wavelengths of 1330 nm and 1550 nm. Despite this apparent monopoly in these devices, other materials are used and do have some advantages over their III-V competitors. Silicon and Germanium are the main rivals to III-V photodetectors and are explored further in this piece of work [1].

1.3 Summary

Photon detection has long been an important part of the communications industry, with interest in a wide range of wavelengths. Avalanche photodiodes are one method of detection, with a more recent focus on detection of weak signals in the infrared region of the spectrum, the limit of which requires detecting single photons. Silicon and Germanium are two materials which may be used to form these diodes. However they are in their relative infancy (around 20 years), as III-V

designs have generally been considered to be the most promising for industrial devices.

This work explores the potential of Silicon and Germanium Single photon avalanche diodes (SPADs), particularly for application to the low energies of infrared photons. It has led to the production of the first successful detection of single photons with a wavelength of 1.55 μm, as well as impressive results at 1.33 μm that compare favourably to those obtained using similar III-V devices. Material characterization and electrical measurements from this work may be used to further advance the designs of future devices.

Work and development into the doping of semiconductors, specifically Silicon and Germanium, via chemical vapour deposition has also been accomplished. Much of the work can be useful in a wide range of growth applications for improving device performance, as well as for initial work into the suspension of doped Silicon boron layers. High doping essentially transforms a semiconductor into a material with metallic properties, which is important for low contact resistance pads for many different devices.

In relation to SPAD devices, there is an interesting possibility for the incorporation of suspended structures which will also be explored in this thesis. While this has the benefit of removing the dislocation networks for mismatched materials, there is also the possibility of reducing the leakage current, as well as introducing a reflective layer on the underside to reflect non-absorbed photons.

Chapter 2

Background Theory

2.1 Basic Properties of Si and Ge

Probably the most important and exploited properties of a semiconductor are its band gap and the mobility of its charge carriers. The band gap is characterized by the energy required to form an electron hole pair in the conduction and valence band respectively. Figure 2.1 shows a 1D representation of the band gap.

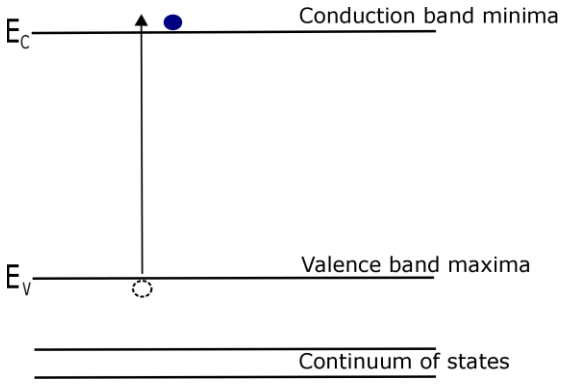


Figure 2.1: Electron excitation leaving a hole in the valence band. The gap between the two bands is the band gap. Energy levels below are core levels.

A semiconductor's band gap may be either direct or indirect. A direct band gap is observed when the maximum of the valence band and the minimum of the conduction band both lie at k = 0. This is the preferred method of excitation as a photon of appropriate energy can excite an electron directly from the valance band to the conduction band. An indirect band gap is seen when the conduction band minima are not aligned with the valance band maximum at k = 0. This instead means that, for a transition to take place, a change of electron momentum is required as well as absorbing the energy of a photon. This momentum can be supplied through a phonon. The two materials used throughout this work are Silicon and Germanium which have band gaps equal to 1.12 eV and 0.67 eV respectively [14], both of which are indirect.

Mobility is of great importance in devices such as transistors, describing how easily carriers are able to traverse through the material. A low mobility of carriers results in slow transistor performance. Both carrier types, electrons and holes, have a unique value for mobility in a given semiconductor material (at a particular temperature). The mobility is given by $\mu = \frac{v}{E}$ where the mobility is μ, the drift velocity is v, and the applied electric field is E. It can be seen that to obtain a

higher drift velocity, and therefore faster transistors, a material with high mobility is preferable. Bulk Germanium has higher electron and hole mobilities when compared to bulk Silicon, which is partly why it has attracted such interest in semiconductor research. However as previously described, the limiting factor which prevents this superior material from surpassing Silicon is its much higher cost.

A crystal lattice is observed where atoms are bonded at regular periodic sites. Both group IV semiconductors, Silicon and Germanium, have a diamond-like lattice structure. This is represented by two face centred cubic lattices which are displaced by a quarter of the diagonal of a cubic unit cell along its diagonal. Each atom in the lattice is bonded to its four nearest neighbours covalently. This results in a tetrahedral system where the atoms are separated by a length described as the lattice constant. Germanium has a larger lattice constant (5.658 Å) than Silicon (5.431 Å) which is important when growing these materials together. For pure Germanium on Silicon there is a 4.2% lattice mismatch which affects the quality of the Germanium layer. It should also be noted that an alloy containing both Silicon and Germanium can be grown ($Si_{1-x}Ge_x$), which will have its own unique lattice constant, defined by the amount of each material (x) in the structure [15].

These alloy structures can also be useful for band gap engineering. Levels of Silicon or Germanium can be altered to tune to a required band gap. This can be useful for photonic applications in which for example, the band gap of Silicon is unsuitable. Interestingly the band gap of SiGe has two different phases depending on the Germanium content. The band gap is Silicon-like up to 85% Germanium, before it becomes Germanium-like. The equations which relate the Germanium content (x) and the band gap are [16]:

$$E_g = (1.155 - 0.43x + 0.0206x^2)eV \quad \text{for } x < 0.85$$

$$E_g = (2.010 - 1.27x)eV \quad \text{for } x > 0.85$$

2.2 Strain enhancement

The mismatch experienced between differing lattice parameters can be exploited when improving device performance. As previously stated Silicon and Germanium have a 4.2% lattice mismatch. When an epilayer of Germanium (Silicon) is grown

pseudomorphically onto a Silicon (Germanium) substrate the lattice parameter of the epilayer matches that of the substrate. This induces strain into the epilayer as the lattice parameter is either compressed or stretched so that the atomic positions align with those in the underlying layer.

When there is this significant build-up of strain between two layers, the most common relief of this is through dislocations. A dislocation is the boundary of a slipped plane. Dislocations [17-19]occur in both the (110) and (111) planes which is useful in diffractive imaging in a transmission electron microscope as stacking faults do not appear in orthogonal diffractive conditions. A misfit dislocation is formed at the interface and is found along the (110) plane, shown in Figure 2.2. Threading arms protrude from the misfit dislocations along the (111) plane and terminate at another surface. The displacement direction, and its magnitude, of the lattice which leads to the formation of threading dislocations is characterized by its burgers vector. Movement of the dislocations occurs along the (111) plane, and is a common method of lowering their density.

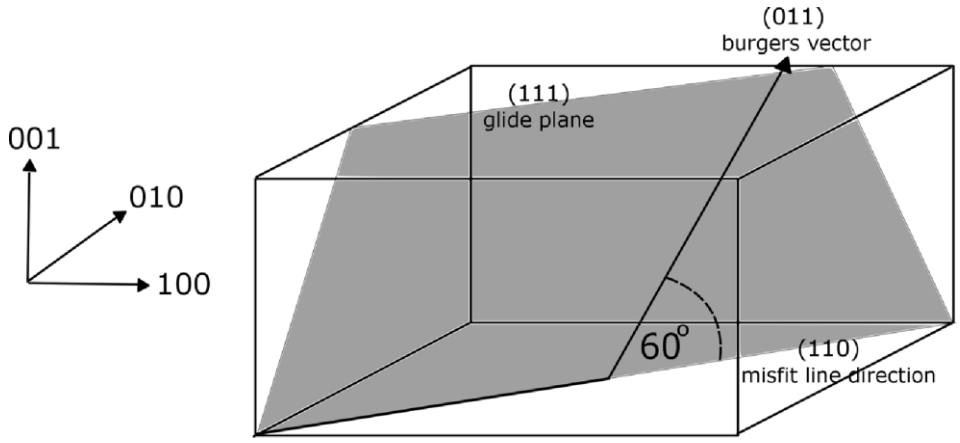


Figure 2.2: Diagrammatic representation of a misfit and threading dislocation. The threading arms direction is defined by the 60^o burgers vector shown in the (011) direction.

As the strain builds up in a layer, the energy increases. At a given value the energy becomes high enough so that a row of covalent bonds may be broken, resulting in the aforementioned slipped plane. At this point the energy exceeds the activation energy required for nucleation of dislocations and relaxation will take place. The point at which a layer undergoes this procedure is known as the critical thickness. Below this limit the layer will be lattice matched to the layer beneath and free of

dislocations. The critical thickness is dependent on both the bulk layer and the lattice matched layer. Large mismatches between layers reduce the critical thickness, and may result in only a few monolayers of growth before the onset of dislocations. The relatively large mismatch between Silicon and Germanium prevents the growth of thick layers (of the order of tens of nm [20]), free of dislocations.

In the case where Germanium is grown directly onto Silicon far beyond the critical thickness there will be a large density of dislocations, particularly around the interface. A characteristic measure of this is termed the threading dislocation density (TDD), and refers to the number of threading dislocations observed in a given area. There are methods of counting, usually directly by plan view TEM or by selective etching to create pits at the dislocation sites, and reducing the number which can be employed to reduce this number. One of which, annealing [21], utilises the potential for the threading dislocations to move and annihilate, therefore reducing the TDD. When subject to suitably high temperatures dislocations are able to glide. Physically the bonds may be reattached at the boundary while others are broken creating an apparent glide of the dislocations. Under these circumstances dislocations may glide all the way to the edge of the structure or they can annihilate when the meet another dislocation. Unfortunately simply annealing a sample to lower the TDD via this method has the drawback of migration of bulk atoms and dopants from their desired location.

Throughout this work the strain in a layer was calculated using:

$$\varepsilon_{perp} = \frac{a_{perp} - a_{SiB}}{a_{Si}} \qquad \text{Equation 2.1}$$

Where the denominator defines what the strain is with respect to. In this case it is the Silicon substrate, but could also be with respect to the bulk material. The perpendicular strain describes that of strain parallel to the growth direction. This equation is specific for a SiB on Silicon system. A negative perpendicular strain value corresponds to a biaxial tensile strain, whereas positive is a compressive strain. For an elastic system this will correspond to an opposite strain value in the in-plane direction i.e. tensile will be positive ($\varepsilon_{perp} \propto -\varepsilon_{par}$).

As the biaxial strain in a layer is increased there is a splitting of the bands. For the valence band there is a splitting of the heavy hole band (HH) and the light hole band (LH) (breaking the degeneracy). For a compressively strained system the HH shifts up and the LH shifts down in energy, relative to their common original energy in the unstrained system. For a tensile strained system the opposite is true. The splitting of the valance bands reduces the scattering between the bands, while the alteration of their curvature (in E-k space) changes the hole effective mass. Meanwhile, the conduction band is shifted to lower energy and, if enough strain is applied, it is possible to change some materials from an indirect to a direct band gap semiconductor. A further result of the reduced scattering rate and reduced effective mass is that the mobility is increased.

2.3 Doping semiconductors

There are two different possible types of dopant for a Group IV semiconductor. These are namely donors and acceptors. A donor will be a Group V material such as phosphorus or arsenic, which have five outer shell electrons. When bound to other Group IV atoms four of these electrons do so covalently. The fifth electron which is left over can occupy a state in the conduction band. This is characterized by a state called a donor level which is present just below the conduction band, and has a much smaller energy required to free it (typically tens of meV compared to ~1 eV for the band gap). An acceptor on the other hand is a Group III atom, typically boron. When this atom forms four covalent bonds with Group IV atoms there will be a missing electron. This represents a hole in the valence band and is characterized by an acceptor level, analogous to the donor level around the conduction band. These extra energy levels are actually only detached from their respective band, shown in Figure 2.3, meaning that at T = 0 K the valence band will still be just full of electrons.

Diffusion in semiconductors is a generally detrimental effect which can lead to performance issues in devices. Essentially it is the migration of atoms from their lattice points through point defects, and may affect both foreign and intrinsic atoms. Dopant diffusion relates to the movement of dopant atoms through the crystal lattice, and self-diffusion describes the movement of bulk lattice atoms.

Both of these types of diffusion are undesirable and impossible to irradiate completely.

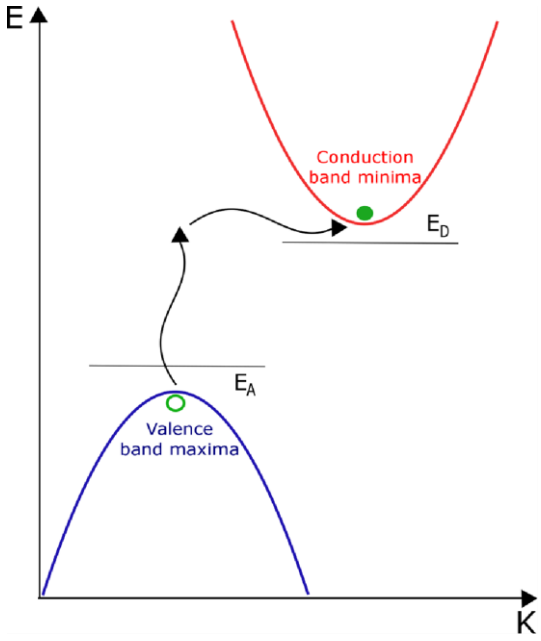


Figure 2.3: Donor and acceptor levels with respect to the conduction and valence bands. The acceptor level is situated just above the valence band maxima, while the donor level is just below the conduction band minima. The blue and red arrows represent the photon and phonon interaction. Along the y axis is energy, and along the x axis is momentum.

There are two main ways in which diffusion is mediated in semiconductors. They are via namely vacancies and self-diffusion. The vacancy mechanism for diffusion, shown in Figure 2.4 is seen where an atom jumps to a neighbouring vacancy site (which is essentially the absence of an atom in the lattice structure). This process may repeat if a vacancy becomes present in a neighbouring lattice site.

For the self-interstitial mechanism, a self-interstitial, which is present in the lattice, will exchange position with a lattice atom so that it is free to migrate through the lattice. This atom will act as a self-interstitial and move through the lattice until it exchanges with another interstitial. For dopant diffusion this is known as the kick out method but is the same as for the bulk self-interstitial mechanism, which is shown is Figure 2.5.

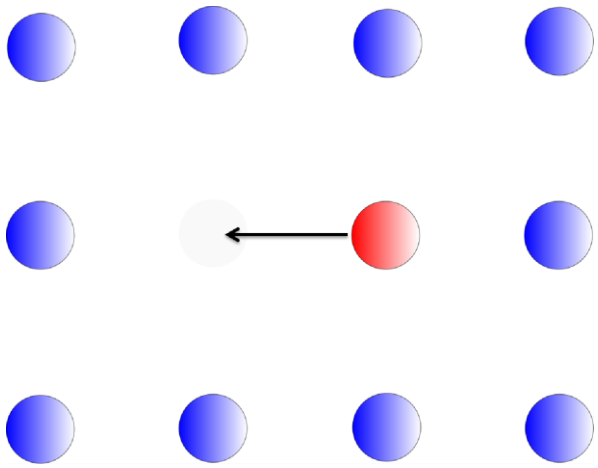


Figure 2.4: Representation of diffusion via the vacancy mechanism. Dopant atom is shown as blue.

The basic equations which describe these two methods are $(AV) \leftrightarrow A_s + V$ and $A_i \leftrightarrow A_s + I$ for vacancy and self-interstitial methods respectively. Here A_i is the interstitial dopant and A_s is the substitutional dopant. The diffusion of atoms is associated with the heat or thermal energy it is given and the diffusion coefficient, which is given by

$$D = K\exp(-\frac{G}{kT}) \qquad \text{Equation 2.2}$$

Where G is the Gibbs free energy, and K is a constant. When G is substituted for

$$\Delta G = \Delta H - T\Delta S \qquad \text{Equation 2.3}$$

The diffusion coefficient can be expressed as

$$D = D_0 \exp\left(-\frac{H}{kT}\right) = D_0 \exp\left(-\frac{E_A}{kT}\right) \qquad \text{Equation 2.4}$$

Where D_0 incorporates the $T\Delta S$ term, and E_A is the activation energy of the dopant atom. Clearly, an increase in temperature in the system gives rise to a larger diffusion coefficient. With a higher diffusion coefficient an atom will diffuse further from its original lattice site. This can be a major problem for precise and abrupt doping profile and interfaces. Typical distances travelled during growth, estimated through Figure 2.6, are of the order of tens of nm, over growth time of tens of minutes. With diffusion tails ideally as low as possible, where doping

concentrations should be close to background over 10-100 nm, it is important to select appropriate growth conditions to maximise the chance of this.

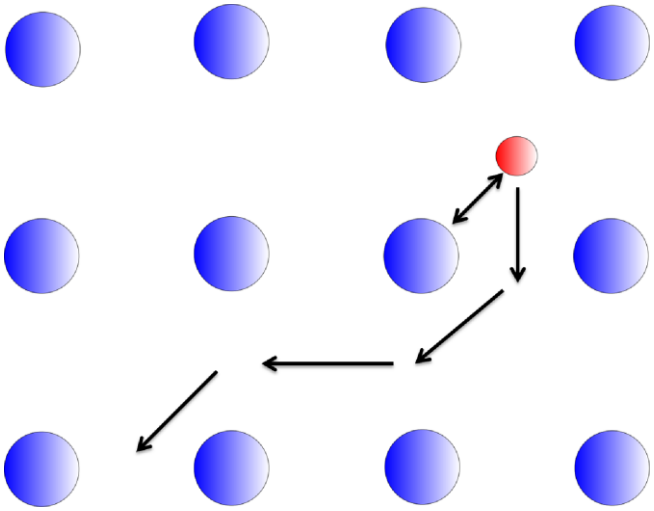


Figure 2.5: Self interstitial or kick out mechanism. The red atom is the interstitial which may swap with dopant or bulk atom.

In Silicon, dopants diffuse through a mixture of self-interstitial and vacancy mechanisms. Boron diffuses through Silicon via the self-interstitial mechanism. This is much faster than n-type diffusion and self-diffusion of Silicon demonstrated in Figure 2.6. There is little data for diffusion in Silicon at temperatures below around 800^{o}C. Therefore the diffusion data of dopants in Silicon presented here is for temperatures higher than those used for this work. However, the general trend can be seen and the level of diffusion in Silicon at room temperature is quite low. At relatively high temperatures used for a long period of time the diffusion may become quite significant.

The diffusion of dopants in Germanium is dependent on which type of doping is present. For n-type dopants the vacancy method is the mechanism by which the atoms diffuse [28]. There is an attractive coulomb interaction between the vacancies and n-type dopants such as P, As, and Sb, which is why this diffusion is faster than self-diffusion of Germanium. Boron is the most researched [29, 30]p-type dopant in Germanium, and is seen to be much slower at diffusing when compared to n-type dopants and even self-diffusion of Germanium, which is shown in Figure 2.7. As there is a low concentration of self-interstitials in Germanium it is predicted that boron diffusion is therefore via the self-interstitial mechanism.

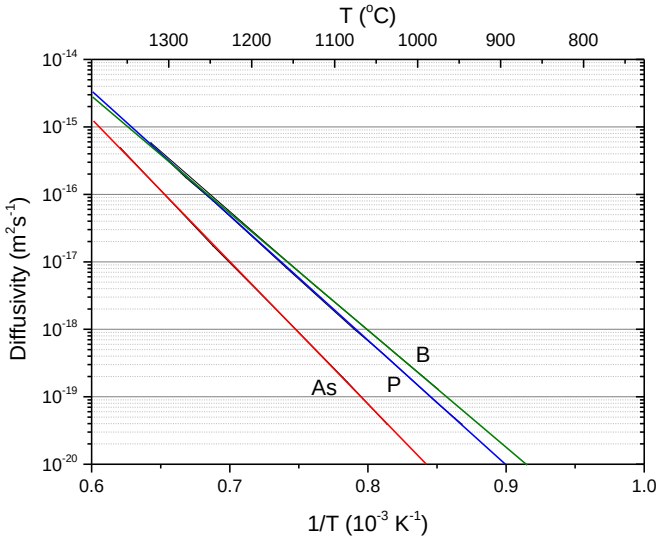


Figure 2.6: Adapted from [27], Diffusivities of common dopants in Silicon against temperature.

There are advantages and disadvantages to each different method of introducing dopants into the material. Implantation is likely to damage the layer of bulk material inducing large numbers of defects. A common way to counter this is to anneal the layer subsequently, which will cause diffusion of the dopants away from the desired area. Diffusing dopants into a layer is difficult to control and is a difficult process when not doping the surface layer. Doping via the epitaxy method does take accurate control of many different growth parameters, which is often challenging, but when performed successfully produces abrupt and accurately doped layers. In particular, when this is done via CVD the structures can easily be reproducibly grown and on a large scale.

In an ideal situation all the dopant atoms may be incorporated into lattice sites and therefore all contribute to the electrical properties of the semiconductor. However in reality each dopant species has a solubility limit, which can prevent efficient doping in an epitaxial layer. These solubility limits in Germanium and Silicon range from $10^{18}cm^{-3}$ to $10^{20}cm^{-3}$[31]. Doping up to this limit allows all the dopant atoms to occupy lattice sites, and therefore be electrically active. Above these limits dopants start to fill non-lattice sites known as interstitial sites. When occupying these sites they are electrically inactive and do not contribute to the

conductivity of a device. Furthermore, when too many atoms are in interstitial sites, and therefore not in the diamond lattice sites, the crystal structure may be disrupted. Clustering of these atoms can result in areas where a different crystal structure begins to form and cause a polycrystalline layer to grow.

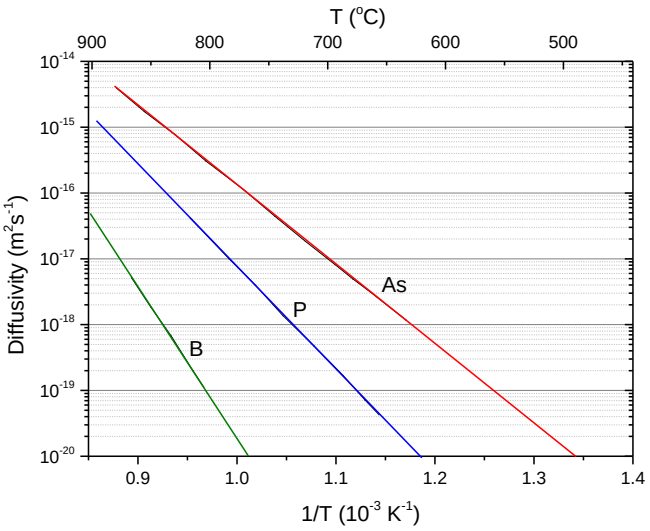


Figure 2.7: Diffusivities of three main dopants used in Ge against temperature, adapted from [27]

2.3.1 Dopant Migration via diffusion and segregation

Ideally a doped layer will retain its profile throughout other treatment post growth. However migration of dopants is often problematic when subjecting structures to elevated temperatures. Procedures commonly use high temperature growth when not doping a layer, or undergo annealing steps to aid the reduction of TDD for example. This increase in temperature leads to diffusion of atoms, whether they are bulk atoms or dopant atoms. Unfortunately this is unavoidable in epitaxial systems as well as other techniques such as implantation. The natural side effect is that either intermixing of layers takes place, for example forming a SiGe layer around a Silicon and Germanium interface, or the doping profile in a structure becomes smeared. Many devices rely on an abrupt and uniform doping profile for operation, which if effected will prevent optimal device operation. For example in an avalanche photodiode abrupt doping profiles are required for a precise electric field in the intrinsic region. If diffusion of the dopants occurs this field may be altered

which could prevent transport of the carriers, give insufficient energy gain for carrier avalanche, or potentially eliminate the field entirely.

2.3.2 High Doping Concentrations

One area of interest for semiconductor application is to dope at such high concentrations that the doped layers begin to exhibit metallic behaviour. This can be used for device contacts, in transistors, and even as an etch blocker. Unfortunately when trying to incorporate high levels of dopants into a bulk crystal such as Germanium or Silicon the quality of the grown layers may be sacrificed. This is mainly due to the issue of passing the solid solubility limit as mentioned in the previous section. In this work I will focus on obtaining high levels of the p-type dopant boron in the Silicon lattice.

Relatively little work has been directed towards highly doped SiB layers, particularly by RP-CVD. Some of the highest Boron concentrations reported are by GS-MBE, up to approximately $10^{21}cm^{-3}$[31]. Other reports have been presented with an electrically active dopant level of around $10^{20}cm^{-3}$ for both SS-MBE and UHV-CVD [32, 33]. The main drawback for producing highly doped SiB samples via MBE time on a mass scale is the increased growth. Although thin layers may be relatively fast to grow in both systems, when including them in other larger and thicker structures the CVD holds a clear advantage from an industry point of view. The solubility limit of boron in Silicon is dependent on the temperature of epitaxial growth. Several reports have shown that the solubility of boron in Silicon is around $8x10^{20}cm^{-3}$ at $1400^{o}C$, $1x10^{20}cm^{-3}$ at $1000^{o}C$, and $2x10^{19}cm^{-3}$ at $700^{o}C$ [34-36]. In this work, most of the growth of doped layers has been at around $700^{o}C$. Higher temperatures such as $1400^{o}C$ are outside the range of RP-CVD and are likely to experience more significant diffusion than at $700^{o}C$ which can be a significant problem for devices that require abrupt profiles.

It is possible to incorporate dopant atoms into a layer at lower temperature ($300/400^{o}C$) by using non-equilibrium techniques, similar to that used for Germanium tin growth [8]. These techniques involve absorbing molecules onto the surface which already have the dopant atom “locked in”, and for this molecule to remain stationary. This technique is often slow and complicated, which is seen in the Germanium tin work. Also novel precursors may be required which can

substantially increase the cost. This work however focuses on growth at higher temperatures (700ºC) under equilibrium growth conditions.

2.4 *pn* junctions

As for many nanoscale devices the basis of a SPAD is a pn junction. These junctions may be used for rapid change of current flow in order to change the state of the device. Fundamentally, a pn junction is simply a region of p doped material adjacent to an n-doped region, but the diode relies on accurately doped p- and n-regions for precise operation. Doping may be performed via implantation, diffusion, or epitaxy as described previously. Initially a pn junction will be described under no bias (no external field), where there will be no current in the junction. The p-region will have a fermi level situated just above the valence band, and the n-region will have a fermi level placed just below the conduction band. Under zero external bias when these materials are joined the fermi level must be uniform across the structure. When the p- and n-regions are brought into contact, free electrons in the n-type material may diffuse across the junction and recombine with a hole on the p-type side. This will create a negative ion and leave a positive charge in the p-region.

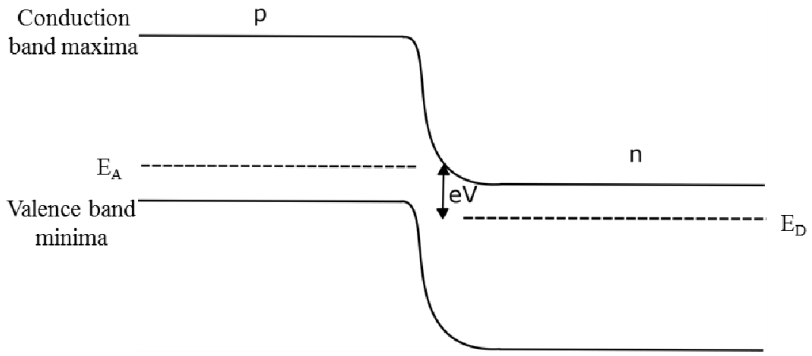


Figure 2.8: Band structure across a pn junction with a reverse bias applied. The applied bias can be seen by the disparity in fermi levels in the p and n regions.

A result of forming the pn junction is a built in voltage. This represents the potential required to move a carrier across the depletion region, and is dependent on the band gap, acceptor level, and donor level. An applied bias can also alter the band gap profile. This is shown in Figure 2.8.

The result of the movement of carriers will be a small internal electric field between the two regions where recombination has taken place, shown in Figure 2.9. This area is known as the depletion region. Electrons and holes are then unable to move across the junction from their respective side due to the new electric field.

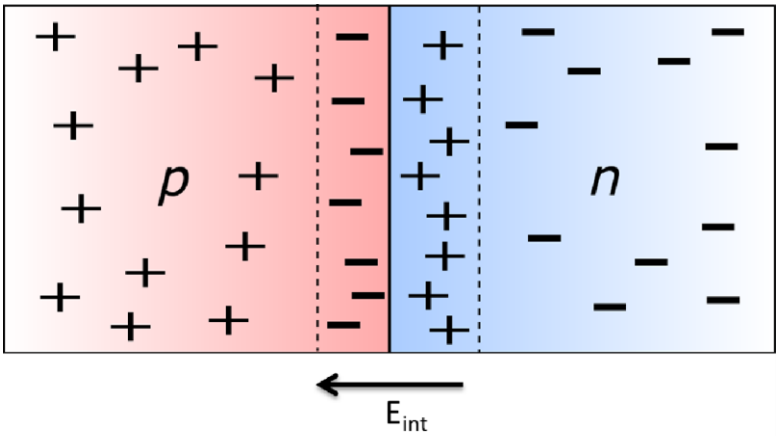


Figure 2.9: Electric field set up in the depletion region of a pn junction.

The peak field is observed when $x = 0$ i.e. at the junction between the two regions. The electric field has a linear relation with distance from the junction. To ensure charge neutrality over the depletion region $N_d W_n = N_a W_p$, shown in Figure 2.10. Therefore if one region is heavily doped relative to the other the majority of the depletion region will lie in the opposite side of the junction.

A pin junction is very similar to the pn junction, with the main difference being a layer of intrinsic material (i) between the pn sandwich. The doping level of this intrinsic region is of the order 10^{15}cm^{-3}. This effectively increases the width of the depletion region and can be very advantageous for photodetector devices. The absorption length of light to be absorbed is $1/\alpha$, where α is the absorption coefficient. This sets a minimum thickness for the intrinsic layer to ensure absorption. The absorption with respect to thickness follows an exponential dependence and varies for different materials [37]. A typical absorption coefficient for Germanium in the near infrared region is in the region of $5\text{x}10^{3}$cm^{-1} which corresponds to an a minimum thickness of around a micron [38]. However thickness much greater than this are detrimental as it delays carrier transport time. Figure 2.11 shows the charge density and the electric field profiles for a pin diode.

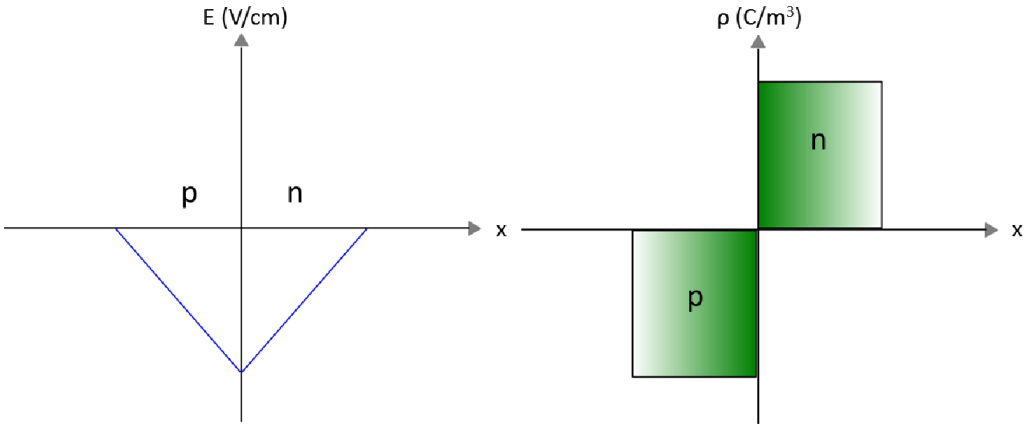


Figure 2.10: Charge density vs position in a pn junction.

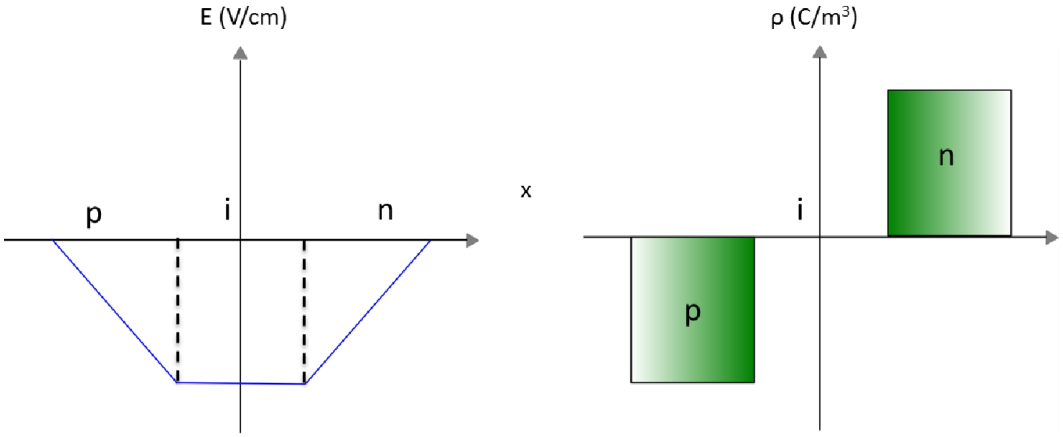


Figure 2.11: pin junction charge density and field with respect to position.

The electric field profile in the pin structure shows the same behaviour for the p and n regions as the pn junction. The electric field across the intrinsic region is constant, and at the same value as the peak field for the pn junction. Pin junctions can be biased in the same way as pn junctions, with reverse biasing being of particular interest for this work. This allows for relatively thick regions of high electric field.

2.4.1 Biasing *pn* junctions

Biasing is an integral part of any device based on a pn junction. There are two possibilities, which are reverse biasing, and forward biasing. By applied a potential so that the field is in the same direction as that of the internal is reverse bias, and applying the field so that it opposes the internal field is forward bias.

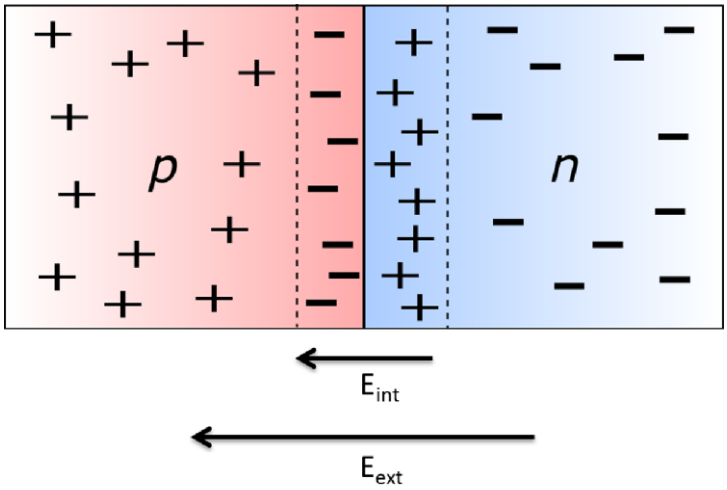


Figure 2.12: A pn junction with reverse biasing, where E ext represents the applied field.

Applications which use forward and reverse biasing for operation include: photodiodes, light emitting diodes, MOSFETS, and solar cells. For a reverse bias system the band gap profile is represented by that described in Figure 2.12. The photodiode is in essence a reverse biased pn junction. Under no illumination no current will flow through the device. However, when photons of appropriate energy are incident on the device electron-hole pairs will be formed. These carriers will be subject to the reverse bias field which will lead to a measureable current. Light emitting diodes (LEDs) are based on a forward bias pn junction. When the device is activated the forward bias is applied and electrons and holes either side of the depletion width are able to recombine and therefore photoluminesce.

2.5 Single Photon Avalanche Diodes

Single photon avalanche diodes (SPADS) are in essence analogous with the photomultiplier tube, which was first proposed in 1930 by Russian physicist L.A. Kubetsky [39]. This device used a primary source of photo-generated electrons and a series of consecutive secondary electron emitters, each with a higher potential than the previous, to produce a huge number electrons. The basic principle is the production of secondary electrons at each emitter, provided that the energy gained through accelerating is sufficiently high to “knock out” another electron. As each event from each electron will produce more electrons the effect is best described as an avalanche. A photocathode is used to produce electrons upon illumination. A

series of dynodes and a final anode are used for multiplication. The emitted electrons will accelerate towards the first dynode and each electron may give rise to one more electron. Each dynode in the PMT will be charged to a higher positive potential than the previous one so after each collision event at a dynode the electrons will be accelerated towards the next one. Multiplication in a PMT is dependent on the number of dynodes and the secondary electron coefficient. Typical values for each respectively are 10 and 4 which give rise to a multiplication factor of 10^6 [40]. The advantage in using a PMT is the large active area available. However they are physically large objects with low single photon detection efficiency [41].

An interesting use of a SPAD is in optical time domain reflectometry. This is used in optical fibres, which often use wavelengths of light similar to those in this work. Photons are injected into the fibre and the detector is required to detect scattered or reflected photons. A SPAD device should be capable of detecting and counting photons at a given wavelength [42].

An avalanche photodiode works in a very similar way, except on a much smaller scale. The simplest design will see a reverse biased pn junction formed from a semiconductor material. More commonly a pin diode would be used so that there is a much larger depletion region. Incoming photons are absorbed and may lead to an electron in the conduction band. Assuming that this electron is in the depletion region it will be subject to the electric field. If the biasing is large enough the energy gained by the electron before a collision will be sufficiently high so that it results in another electron in the conduction band. This is known as impact ionization, and the repetition of these events via the primary and secondary carriers leads to an avalanche current depicted in figure 2.13. Gains of several hundred are common in avalanche photodiodes. It is also important to have a high enough bias so that the intrinsic region is fully depleted. The multiplication/gain for such as device can be described using the following equation:

$$M = \frac{1}{1-\int_0^L \alpha(x)dx} \qquad \text{Equation 2.5}$$

It can be seen that an increased impact ionization coefficient (, or β for holes) for a given material leads to an increase in the gain for the device. Larger gains are

beneficial for detecting a photogenerated signal. The excess noise factor, which describes the noise associated with the multiplication process, is calculated through:

$$F = kM_e + (2 - \frac{1}{M_e})(1 - k) \quad \text{Equation 2.6}$$

$$F = \frac{1}{k}M_h + (2 - \frac{1}{M_h})(1 - \frac{1}{k}) \quad \text{Equation 2.7}$$

Where

$$k = \frac{\beta}{\alpha} \quad \text{Equation 2.8}$$

It can be observed that the excess noise factor for electrons and holes (M_e and M_h respectively) oppose each other, i.e. multiplication via one carrier will give rise to a large amount of noise relative to the other. This makes the choice of material important, particularly where multiple materials are used. The excess noise factor can also be applied to SPAD devices.

Design of the layers is crucial in producing efficient photon detection devices. Generally a thick intrinsic layer will increase the performance of such a device. This will increase the likelihood of the absorption of photons as there is a larger amount of material to pass through.

The absorption coefficient (α) and its dependence with the wavelength (λ) of light to be absorbed are expressed through the following:

$$\alpha = \frac{4\pi k}{\lambda} \quad \text{Equation 2.9}$$

The absorption coefficient is inversely proportional to the thickness required for absorption.

Unfortunately, simply growing a very thick intrinsic layer is not the ideal solution. A drawback to thick layers is that the transit time of any carriers will be increased, overall slowing down the device. Therefore a trade-off must be met between the two opposing factors which will be dependent on the materials used and its application. It should also be mentioned that some carriers may be produced

outside of the depleted intrinsic region. Instead of being subject to a high electric field the carriers will have to rely on diffusion which is a significantly slower process than drift via the electric field. It is therefore sensible to reduce the thickness of any doped regions as much as possible.

A SPAD is not very dissimilar from the avalanche photodiode. While the multiplication gain for an APD can be quantifiable, the gain for an SPAD should theoretically be infinite. The other main difference is that a SPAD is under a much larger reverse bias. Under operating conditions this will be above the breakdown voltage and is known as geiger mode. In the case of a SPAD a single photon will be incident on the intrinsic region giving rise to an electron (and hole). It should be noted that either carrier (electron and hole) may be used for device operation, but as the SPAD in this work uses electrons, electrons will be described throughout this theoretical explanation. Under the larger reverse bias the electron will repeatedly undergo impact ionization, as will each secondary electron which leads to an avalanche breakdown of the device. The excess noise factor associated with the device is consistent with equations 2.6 and 2.7, which is dependent on the type of carrier which initiated the avalanche.

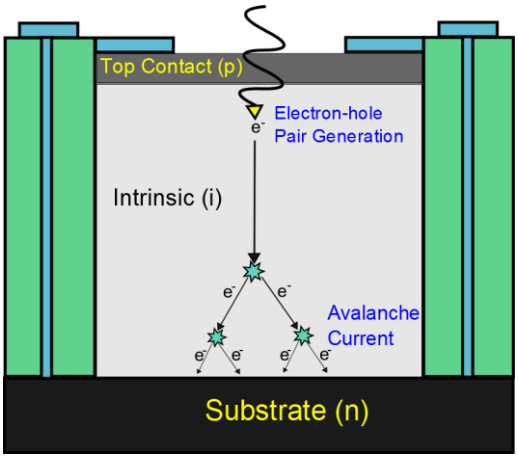


Figure 2.13: Schematic of the avalanche process in an APD. Incident photon is absorbed in the intrinsic region where it is subject to a high electric field, leading to breakdown through impact ionization.

There are two main factors which define the capability of a SPAD to detect and produce a signal. These are namely the detection efficiency and the quantum efficiency. The quantum efficiency describes how likely an incoming photon is to

produce a measureable current, whereas the detection efficiency defines the likelihood of photon absorption resulting in an electron hole pair. It will depend on factors such as the growth quality, doping profile, reverse bias, material choice, and the wavelength of choice. As described earlier in the chapter the process of absorption of a photon is subject to the type of band gap of the semiconductor. The most popular group IV choices have indirect bandgaps which are less suitable for photon absorption. However the advantage of using Germanium and especially Silicon, is the low cost in production and the obvious compatibility when integrating these devices onto other platforms which will generally also be Silicon based. The other option which is commonly used for SPADs is from III-V materials. These materials are often more efficient in detection of infrared wavelengths than their group IV counterparts and are a popular choice [43]. A significant drawback to the use of III-V materials for SPADs is their incompatibility with a Silicon dominated circuit. The compromise of lattice matching III-Vs to Silicon involves growing a SiGe buffer which is defective and often not desirable for SPADs [44].

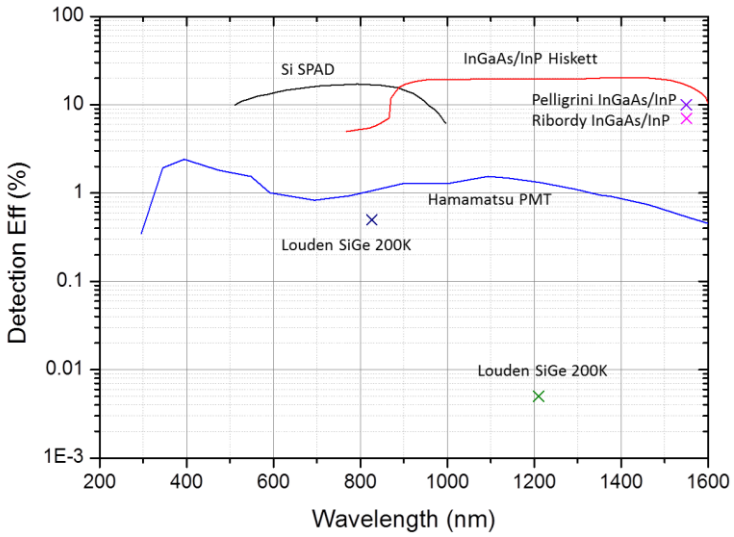


Figure 2.14: Plot demonstrating detection efficiency of several different SPAD devices of varying material design. [43]

A summary of SPAD detection efficiencies for a selection of prominent materials is presented in Figure 2.14. It is evident that the most successful SPAD devices in the range of 1330 nm and 1550 nm are those developed using InGaAs/InP. Very

few attempts have been made with Group IV materials. A detection measurement of 10^{-6} (10^{-4}%) for a Ge and Si based design has been made, but was not represented on this graph [45]. The potential for Silicon, which is represented by the solid line does demonstrate the suitability for devices operating at the shortest infrared wavelengths and even the visible part of the spectrum. As can be observed in Figure 2.14, SPAD measurements at wavelengths above 1330 nm are rare, with no measurements approaching the level obtained with III-V materials. The lack of comparably efficient devices has been caused by the low quality of Germanium layers. The improvement of this is one of the keys to this work.

The two main options for growing SPAD structures are through vertical or planar orientations. For planar devices doped regions are formed through implantation in the areas required, whereas the vertical structures are capable of using dopant incorporation during the growth [46]. In this latter approach it may be difficult to accurately dope the structures to ensure the correct electric field profile. The work carried out in this thesis focuses on vertical CVD growth so that dopant positions could be optimised and reproducible.

2.5.1 Dark Counts

A significantly detrimental factor in the device performance of a SPAD is the dark count. This major problem seriously limits the usability by producing false counts, i.e. non-photogenerated counts. Measurements of this are performed under no illumination from any source. There are three source of dark counts which are, thermally generated carriers, trapped carrier (which is discussed in section 2.6.2), and tunnelling. At low temperatures there are no free carriers in the conduction band, but as the temperature is increased the population in this band will slowly increase. These thermally generated carriers may undergo impact ionization in exactly the same way as those produced via a photon. Under operating conditions it is impossible to decipher the initial cause of an avalanche due to these carriers. A solution to this is to operate only at very low temperatures (77 K) which prevent these dark counts. However the band gap of a material will decrease with temperature which directly leads to an increase in detection efficiency for a given wavelength. Also it may not be ideal or practical to operate such a device at such a

low temperature. Nevertheless, knowing the dark count rate at these temperatures is useful for any device to estimate the number of thermally generated dark counts.

Detection efficiency increases with increasing excess bias voltage. This is caused mostly through an increased chance of an avalanche being triggered. Unfortunately with the increase in excess bias voltage there is also an increase in dark counts. This is due to the increased chance of avalanche in the same way as for the detection efficiency as well as field enhanced dark counts [47].

Another effect which can be detrimental to the operation of a SPAD device is spacial tunnelling. This is observed when the field in the absorber region is too high. At high fields it is possible for carriers to tunnel from the valence band in the p region across the potential into the conduction band of the n region. With this tunnelling effect avalanches can be initiated without an impinging photon with sufficiently high field (Zener breakdown). The tunnelling current density is proportional to the field and band gap according to the following equation.

$$\frac{I}{A} \alpha \exp\left(\frac{-\varepsilon_g^{\frac{3}{2}}}{E}\right) \qquad \text{Equation 2.10}$$

I/A is the current density, ε is the band gap of the material, and E is the field applied. Conditions where a higher electric field is applied and/or a small band gap material is used will lead to an increase in tunnelling current, therefore increased overall dark current for the device.

To reduce the risk of this it is important to keep the electric field high in the multiplication region, and much lower in the absorber, while still allowing for carrier transport. Appropriate doping throughout the structure helps to determine the electric field profile. In this work (section 4.4.5) an intermediate layer is used to tailor the electric field in each region. While an increased band gap should also reduce the tunnelling current, the upper limit of the wavelength which can be absorbed by the material is reduced. Therefore this may be unsuitable for applications requiring, for example, IR absorption or longer.

2.5.2 Afterpulsing and Device operation

As mentioned above afterpulsing is one of the main causes for dark counts. This phenomenon limits how often a single photon can be detected. Dislocations, defects, and interfaces which may be present in a layer, formed during growth, are able to trap carriers as they traverse the diode. While dislocation and defect densities may be reduced through various different techniques, such as annealing or two step growth, there will always be a small level present (unless the device is a homostructure) and therefore avoiding afterpulsing completely is impossible. The carriers may fall into one of the associated trapping states while still in the high field region for a time t. If the device is biased above breakdown when the trapped carrier is released it may undergo impact ionization, much like in the case of a thermally generated carrier. The standard approach to accommodate this problem is to pulse the bias above breakdown. This is called gated operation and is presented in Figure 2.15. This means that the device will only be operational when the bias is taken above the breakdown. During the time where the bias is below breakdown (dead periods) any trapped carriers released will not undergo impact ionization and will not lead to avalanche breakdown. Instead they will be swept out of the depletion region. The dead period for a given device should allow all the trapped carriers to be released without risk of a dark count. The afterpulse probability is an exponential function and the dead period should be at least as long as the trapped carrier lifetime. Clearly long dead periods are not suitable for operation and should be minimised. As dislocations and defects are a source of trapping states it is paramount to reduce the level of these as much as possible. Typical pulses are of the order of a few nanoseconds, and the dead period may be a few microseconds [48].

The other drawback to gated operation is that the arrival of a photon needs to be known with a certain degree of accuracy, and therefore limits the efficiency of detection particularly when it is not known. However under non-gated operation (constant bias above breakdown) the devices results are almost meaningless as it is impossible to know if a count was dark or generated by a photon.

High threading dislocation densities can lead to an increase in the leakage current in a device [49]. Carriers may be trapped at sites of dislocations before releasing at

a later time as a leakage current. This highlights the importance of minimising threading dislocation density (TDD) in and around a heterojunction when optimising a device.

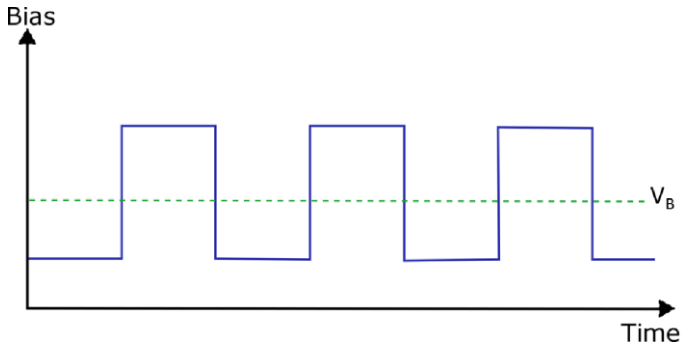


Figure 2.15: Example of pulsing the Bias above breakdown for short periods. Precise over bias and dead times depend on the device and its application.

The over-bias used is also an important factor. While maximising the overall chance of detection it is possible to simply increase the bias voltage. However, with this leads to a larger amount of charge flowing through the device and hence more filled traps. Subsequently longer dead times are required for devices operated in such a way which is a drawback [50].

2.5.3 Single Photon Avalanche Diodes designs (Si and Ge or III-V)

Arguably the most popular choice for a SPAD, particularly for detection of infrared radiation, is an Indium-phosphide based design. This design also introduces the concept of separate absorption and multiplication (SAM) regions. Unlike earlier avalanche photodiodes where the detection occurred in one intrinsic region of a *pin* diode different materials may be selected for each stage of the detection process. The absorption will take place in one intrinsic region while the multiplication will take place in a separate intrinsic region. There are two different methods in which this can be employed depending on the preferred carrier type. Generally for electron induced avalanches a *pipin* structure is used, and for hole induced avalanches a *ninip* structure is used. For each design there will be a different electric field profile in each region with the general concept shown in figure 2.16. The absorption region will experience a relatively small electric field, and the multiplication region will experience a high electric field. The low electric field in the absorber region is to prevent any impact ionization events and is used to carry

the electron or hole into the multiplication region. The high field in the multiplication region should be above the breakdown voltage for that material to ensure an avalanche is possible. A thin lightly doped region known as the charge sheet is grown in between the absorber and multiplication region to coordinate the field in each section.

Under a *pipin* structure, depicted in Figure 2.17, photons will be absorbed in the first intrinsic region between *pip* where the p-doping in the middle in sufficiently small to allow punch-through of the electric field through from the *pin* region. The electrons will then drift into the intrinsic region below, while the hole will drift up towards the top p-contact. Punch-through is characteristic of when the depletion region stretches into the pip region where a small field will exist and there will be efficient transfer of carriers between the two regions. Therefore it is important that punch-through is reached before breakdown.

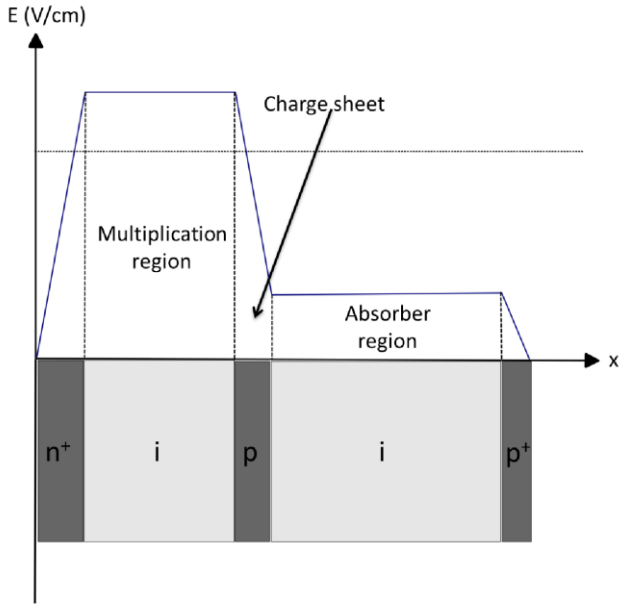


Figure 2.16: Electric Field profile over a separate absorption and multiplication region SPAD. Field is above breakdown in the multiplier and below in the absorber. SPAD structure is presented below for visual representation. For SPAD operations the electron will drift from the absorber to the multiplication region.

The punch-through voltage is generally considered independent of the temperature of the device. However, breakdown may occur at a lower voltage at lower

temperatures due to a reduction in scattering events from phonons which may prevent sufficiently high energies of the carriers to be reached. A higher breakdown voltage is beneficial in preventing breakdown before punch-through, which gives Silicon an advantage over Germanium as a multiplication region material due to its larger breakdown field. Also Silicon has more desirable multiplication statistics than Germanium which is why it is important to include it in a Group IV SPAD. Multiplication in a device is higher when the two impact ionization coefficients for electrons (α) and holes (β) differ. For Silicon α is a lot larger than β (a factor of around ten times[51, 52]). For Germanium both values are very similar which is why it is not suitable for the multiplication region [37, 53]. The ratio between these two factors describes the excess noise factor which is important for measuring the signal, and therefore the reason for Silicon being preferred over Germanium for the multiplication region. As well as the high electric field it is important that the thickness of the region is much larger than the mean free path of the carriers. Typically this should be of the order of microns, which is due to the inverse proportionality with respect to the impact ionization coefficient, while the mean free path of electrons in Silicon is only around 10nm [54].

The importance of punch-through is highlighted when comparing two commercial devices (InGaAs/InP) where only one had achieved punch-through [55]. In the case of the SPAD which had not punched-through the detection efficiency was effectively zero, and only became functional when the temperature was increased sufficiently to allow punch-through.

Once the electron reaches the multiplication region it will be subject to the high electric field and may give rise to an avalanche current.

The *ninip* design works in much the same way. In this case the absorption will again take place in the absorption region which is the *nin* region. In this case, the electron will drift up to the top contact and the hole will drift into the *pin* region. Once the hole reaches this high field region it may undergo impact ionization in the same way as with the electron system. The choice of design is often decided by the materials used, as the impact ionization coefficient for electrons and holes is likely to be different and therefore be favourable in one way to increase the chance of avalanche breakdown.

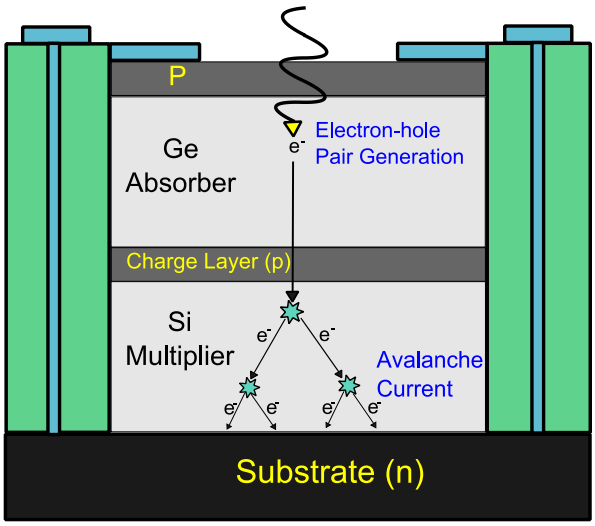


Figure 2.17: Schematic of a SiGe separate absorption and multiplication SPAD.

The indium-phosphide (InP) based SPADs use the *ninip* design as they utilise holes for avalanche currents. The multiplication region of such devices is formed from InP where a high electric field exists, and the absorption of photons takes place in a InGaAs absorber region where the field is low.

The first Group IV SPAD device was made using only Silicon. The cheapness and compatibility of the material made it a convenient choice. Such devices have the capability for detection in the spectral range of 400-1000nm. Wavelengths beyond this limit, such as infrared, are incapable of being detected due to the low energy of the photons. This renders Silicon only SPADs effectively useless for applications which require absorption of such wavelengths. Despite this detection efficiencies for single photon detection have peaked at well over 60% in some designs for wavelengths up to around 800nm, where very thick layers were used to increase the probability of absorption [56]. A drawback with these devices was that the time taken between the arrival of a photon and count being measured (jitter) was relatively long (1ns), due to longer transit times of the carriers.

Despite the unsuitability of Silicon for single photon detection for infrared radiation it could still have great value for the next generation of group IV SPADs. Silicon has a high breakdown field ($\approx 3 \times 10^5$ V/cm) as well as a high impact ionization coefficient making it ideal for a multiplication region. Its favourable impact ionization coefficient allows it to be preferred to other materials more

suitable for IR detection, despite its relatively poor ability to detect above 1 μm [57]. The introduction of Germanium opened up the potential for absorption of infrared light. The first APD was proposed in the mid-1980s which demonstrated absorption of infrared light (λ=1.3 μm) in a SiGe absorbing region [58]. It wasn't until 2002 that the first SPAD was designed and tested using both Silicon and Germanium [59]. Multiplication again occurred in the Silicon region and absorption in the SiGe. A series of multilayers of $Si_{0.7}Ge_{0.3}$ were used for the absorption region alternated with pure Silicon layers. These strained SiGe layers allowed absorption of single photons with wavelengths of 826 nm and 1210 nm. It was also important to keep the layers below the critical thickness to reduce the defects and therefore dark counts. Unfortunately this also reduced the likelihood of absorption which contributed to a reduction of detection efficiency. Detection efficiencies of 0.1% and 1% were obtained for 826 nm and 1210 nm respectively, with a dark count rate of 10^7 counts per second (cps). Although these efficiencies were very low they were able to highlight the potential for Si/Ge SPADs as these devices showed over an order of magnitude improvement compared to an all Silicon SPAD [59]. Despite reporting detection of single photons, $Si_{0.7}Ge_{0.3}$ is far from ideal for detection at wavelengths above 1μm. With only 30% Germanium the band gap is likely to be just over 1 eV, whereas to obtain a small enough energy gap for absorption of photons of 1.55 μm the Germanium content should be over 85%. The reason these SiGe layers were used was due to the poor quality of pure Germanium produced via any crystal growth technique at the time.

As Germanium growth has improved in the last decade or so Si/Ge separate absorption and multiplication SPADs have been realised, allowing for a more serious attempt at producing efficient devices. Germanium growth still involves levels of dislocations, which are detrimental for device performance, and has attracted various methods to reduce the TDD. Introduction of SiGe buffer layers has been proposed to reduce the defects in the absorber region of Ge on Si photodetectors [60]. However, this thick SiGe region may not be ideal for all devices if carriers must transit this region, since there is still a large TDD throughout these layers.

An alternative approach which attempts to avoid trapping from highly defective Germanium layers is to position the electric field outside of areas with a large TDD

[45]. A highly doped Germanium layer is grown on a thick n doped Silicon layer. This ensured that the depletion region was mostly in the superior Silicon layer and that the electric field was shifted away from the Germanium/Silicon interface. Despite the extremely high TDD in the Germanium this should reduce the trapping effects and reduce the overall dark count of the structure. This did provide a relatively low dark count rate but gave very low detection efficiencies at infrared wavelengths, due to the majority of the absorption occurring in the Silicon and the increase in trapping through defects.

The main competitor to these group IV SPADs is those designed using III-V materials, particularly InGaAs and InP. These devices have consistently performed to a higher detection efficiency, especially around the 1550 nm wavelength, and are often regarded as the benchmark for other devices [2]. The narrow band gap of InGaAs makes it suitable for absorption of infrared photons and InP, with its much larger bandgap, is used as the multiplication region. Much like the Silicon and Germanium devices, the SAM concept is commonly used for InGaAs/InP SPADs. Two main drawbacks to using this material are the difficulty in integrating with other Silicon based devices, and the offset in the valence bands between InGaAs and InP which produces significant problems in the devices. The most common technique is to use grading layers between the two materials. This reduces the effects of trapping which is prevalent across the interface [50].

Work has been performed by Pelligrini on devices with either one or three intermediate layers separating the InGaAs and InP. By introducing three steps to the grading the holes were more likely to drift across the interface without becoming trapped, illustrated in Figure 2.18. The discontinuity is more severe if only one intermediate layer is used, and is particularly detrimental to the DCR. Detection efficiencies reached 10% at a wavelength of 1550 nm comparing well to the best reported[61]. The highest reported SPDE for a III-V device in the infrared range has been 45% at 1310 nm. This device was designed using a 2 μm InGaAs absorber layer (with 53% In), and a 1 μm multiplication layer. The thickness of these layers gives an indication for what should be suitable for a SPAD device. This device was operated in gated mode with a pulse of 4 ns above bias. It was found that the majority of the dark count came from leakage currents, which would fill up a significant number of traps. To counter this problem a DC bias was applied

(just below breakdown) to remove as many trapped carriers as possible while the device was not active [62].

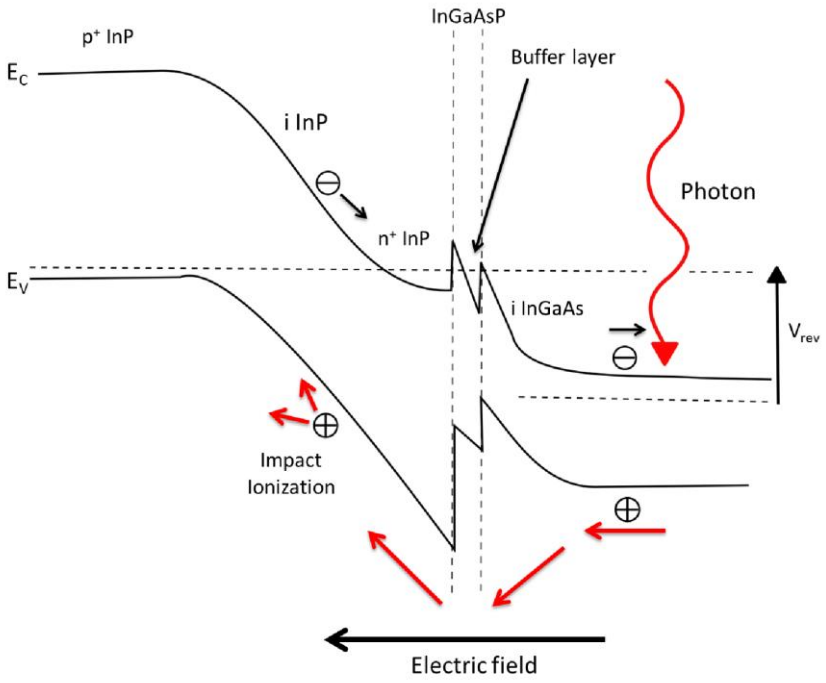


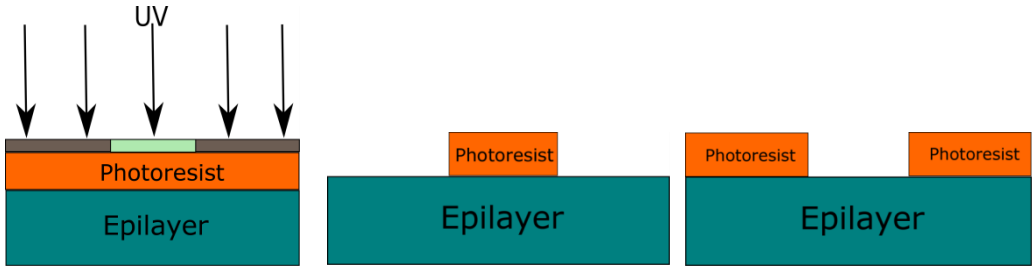
Figure 2.18: Band diagram for III-V SAM-SPAD with buffer layer used to overcome band gap discontinuity issue. Photon is absorbed in the InGaAs before traversing the buffer and undergoing impact ionization in the InP layer.

The detection efficiencies of InGaAs/InP SPADs have led the way in SPAD technology but the need for intermediate layers does present a flaw in their functionality. Effects of trapping, seen even with designs using multiple intermediate layers, require the use of gated mode operation. In InGaAs/InP SPADs trapping times can be relatively long, thus requiring long dead times which can be around 100 μs. This potentially low repetition rate (how often an overbias gate occurs) can make it unsuitable for some applications, which provides an advantage for Silicon and Germanium based devices.

2.6 Photolithography Processes

Patterning the surfaces of epitaxial structures is a vital process in device fabrication. While this can be utilised for devices such as SPADs, it can also be useful for manipulating structures before further processing [63]. This is the case for suspended structures such as membranes and wires. Both structures were used

during this work, and were designed using the photolithography process. The general idea is that a pattern can be imposed onto a surface to protect specific areas from any further etching processes [64]. This can allow for these unwanted or unnecessary areas to be removed without damage to the whole structure. A key part to this process is the use of photoresist and UV radiation.



(a) (b) (c)

Figure 2.19: Diagrammatic representation of the photolithography process. Image (a) shows the general photolithography set up. Image (b) shows negative resist, and (c) shows positive resist.

Photoresist is used as a protective layer on the surface of a sample. Generally a mask is used to pattern the area covered by the photoresist resembling the device which is to be fabricated. This mask is used to protect regions of the deposited resist from UV light during the process. There are two types of photoresist, either positive or negative. For positive resist when the layer is exposed to UV radiation the photoresist changes its chemical structure so that it becomes more soluble to a developer solution. Therefore when the sample is subject to this developer the areas which have been exposed to the light will be removed resulting in an area void of the resist. For a negative resist the opposite occurs. When a negative resist is exposed to the UV radiation the resist becomes harder to dissolve, as it is polymerised. The results is that when the sample is developed in the solution the areas protected by the mask are dissolved but those areas which were exposed are not. This process is demonstrated in Figure 2.19. For the suspended structure designs the areas which were to be removed were not protected by the photoresist, i.e. a positive resist was used. Material in between the areas with protect were etched by the solution, which was resisted by the photoresist layer.

2.7 Suspended Membranes and Wires

2.7.1 Etching

The etching process is used to most commonly selectively remove material from a structure. Photolithography may be used to pattern a surface with areas which may be resistant and other areas which are not. When subject to an appropriate etchant the areas which are not resistant will begin to etch away. The time the sample is submerged in the solution dictates how much material is removed. Despite removing more material with more time, the process is not always linear and should be calibrated first.

Some materials will be resistant, or have extremely slow etch rates, to certain etchants which is useful for stopping the process. Etch rates may not accurately be known, so it is often useful have a layer of material which will halt any further etching. An example of this is the resistance of Germanium to a TMAH etchant [65], which allows a region of Si substrate to be etched away over an imprecise time leaving a Ge over-layer unaffected (see Figure 2.20).

Etching may also be performed to estimate the TDD in a sample. Threading arms which reach the surface of a structure etch at a faster rate than the bulk surface material, so selectively etching a sample creates craters, or etch pits, locally around the threading arms. Generally each crater is assumed to be one threading arm, and by counting the number of pits in a given area a value for TDD can be gained. The actual etchant used depends on the surface layer, with different HF based solutions used for etching Germanium and Silicon due to the different etch resistance of the materials. It should be emphasised that this method is only an estimate of the TDD. The pits do not necessarily represent only one threading arm as they may have merged, and is only representative of the specific area investigated. It is a useful technique for showing whether or not the TDD is low, but absolute values hold a large error. Samples in this work were etched and five different regions were studied to give a value of TDD. For etching Ge an iodine solution had to be used in a ratio of 5:10:11 ($HF:HNO_3:CH_3COOH$) and diluted with DI water in the ratio 3:1 (due to the iodine etchants aggressive etch rate).

2.7.2 Membrane etching

The etchants used in this work were either TMAH or an HF solution. TMAH etches along all planes except for along the (111) planes, due to a high density of packing. The resultant etch pit is shown is Figure 2.20, where a distinct V shaped characteristic is observed.

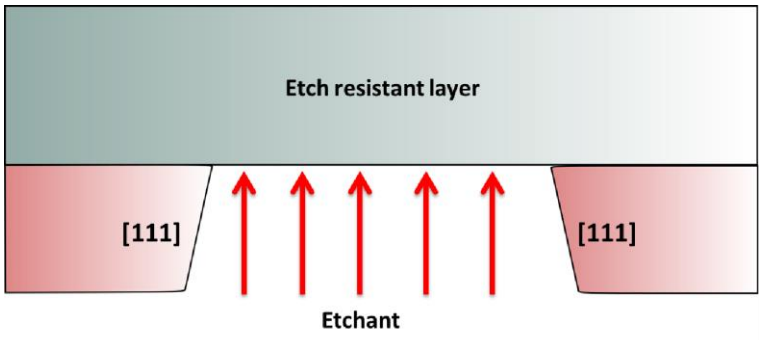


Figure 2.20: Diagram of etching at planes. The substrate (red) is etched away by the TMAH solution leaving the etch resistant epilayer (grey).

Assuming that the layer succeeding the substrate is at least etch resistant it is possible to suspend small squares of epilayers. One interesting use for this is the potential removal of the misfit dislocation network from a Germanium layer once the Silicon has been etched away [65].

For this work the material in question is SiB. Although pure Silicon is not resistant to the TMAH, when high enough levels of dopant atoms such as boron are introduced into the epitaxy, etch resistive behaviour becomes apparent [66] [67].

2.7.3 Suspended works

Suspension of epitaxial structures has become a novel and interesting technique in recent times. An advantage of this is that the effects of the substrate can be removed and essentially make the suspended structure electrically, thermally, or structurally independent of its substrate. Prior to any etching techniques the substrate would be removed using a machining method [68].

A relatively expensive method used for fabrication of suspended structures is that of dry etching. Cost is generally accumulated through processing steps which often require high temperatures. The cheaper option is to utilise wet etching stages

instead. Generally structures are grown onto substrates which have no importance to the device operation, and therefore are expendable. The standard etchants used for this area of work are potassium hydroxide (KOH) and tetramethylammonium hydroxide (TMAH). These chemicals can be used for both backside removal and a top down underetch approach (useful for making bridge structures).

It is of great importance that etch-masks or etch stops are selected appropriately, as the production of suspended structures depends on only certain areas having exposure to the etchant. It is common for insulating layers, particularly SiO_2 or Si_3N_4, to be used as resistant layers to the etchant. Similarly doped layers can often present much slower etch rates than the bulk Silicon, especially with increased doping levels [69]. Li *et al* took advantage of the very low etch rate of polycrystalline Germanium (to KOH and TMAH) to use it as a mask. They were also able to produce a Germanium membrane, although the quality was such that any further application was not possible [70].

Electrochemical etching has been used to design n-doped membranes of thickness between 3 and 8 μm. Doping of the layers was performed using a bubbling flask to obtain concentrations around $10^{17}cm^{-3}$. An oxide mask was used on the backside to establish a series of squares (between 5 and 0.1 mm) which could be etched to create the membranes. The p-type substrate used as the initial growth platform was unresistive to the KOH etchant in the areas which were not covered by the oxide, whereas the n-type material was [71].

An interesting investigation explored the removal of a substrate with epitaxial Germanium on top. As with mismatched materials high levels of dislocations are often present which can hamper the electrical capability of any fabricated device. By removal of the Si/Ge interface it was proposed that the dislocation network could be removed through instigating glide. The thickness of these structures was 200 nm, which demonstrated the suitability of Germanium as an etch stop. Van der Pauw measurements showed an improvement to the electrical isolation of the bridges when compared to the bulk material full of dislocations. It has been proposed that these structures could have applications for Ge lasing [72] as the strain imposed on the layer could be enough to obtain a direct bandgap. Application of around 2% tensile strain is generally thought to be enough to shift

the conduction band to a direct band gap position, which is necessary for lasing [73, 74].

With leakage currents through dislocations a common problem in Germanium based devices; the removal of this network should form better quality Germanium and therefore improvements for device application. A recent study on Germanium membranes described a technique of growing compressive and tensile strained Germanium on a double sided wafer. The tensile strained layer was produced by using the cooling process and the larger thermal expansion coefficient of Germanium which does not contract as much as the Silicon substrate. The highly defective compressive layer was found to be useful as a backside mask, while the higher quality tensile layer was seen to be an ideal platform for further growth on top of the membrane [75].

Another Germanium membrane has also been produced by Nam et al. in which the applications for Germanium lasing has been identified. These relatively thick layers (1.6μm) have imposed tensile strain upon them due to the cooling process and thermal mismatch between the Silicon and Germanium. With this strain Germanium can theoretically be converted to a direct band gap. Normally injected electrons will sit in the lower L valley, and therefore be unable to contribute to a radiative response, but with a strain of nearly 2% this does not present a problem. For the mask SiO_2 was used to cover the backside with a pattern, while the etching was performed in a TMAH bath at 90^{o}C for 6 hours. This resulted in a suspended tensile strained Germanium membrane [64].

Creating etch resistant masks for epitaxial structures can be problematic due to the high temperatures which can be required for their production. Particularly high temperatures can cause serious amounts of diffusion in the structure, which when suspended has a different composition to that originally specified. This makes a low temperature mask an appealing option when dealing with heterostructures.

While these examples explore suspended structures that are relatively small in size, there has been work investigating the potential of completely suspending an epilayer from its substrate. Roberts et al. grew a thin SiGe/Si sandwich on an SOI substrate initially, before bathing in an HF solution. This selectively etched away the buried oxide layer leaving only the Si/SiGe/Si sandwich. Upon removal of the

oxide the slightly compressively strained SiGe was able to relax, therefore imparting a slight tensile strain of the Si, which could be used for further growth or device application [76].

While many different materials have been explored for suspension, little work can be found regarding SiB. In general the applications using bulk Group IV materials and similar revolve around the strain implications of suspending such structures, and how the localised areas of strain can be utilized. While the strain and tilt is investigated in this work, believed to be for the first time, the application with regards to SPAD devices presents an interesting route for device optimization. This follows in the later chapters, where pioneering work is laid out to move towards a suspended SPAD.

Chapter 3

Experimental Techniques

3.1 Transmission electron microscopy

Structural analysis in this work was performed using a JEOL 2000FX Transmission electron microscope (TEM), at an operating voltage of 200 kV. This microscope uses a W/LaB6 filament as the electron source and has a GATAN ORIUS 11 megapixel digital camera to record images. This TEM allowed for magnifications between 1000 and 200,000 times, which was more than sufficient for the samples studied in this work.

Transmission electron microscopes are an invaluable tool for characterizing a multitude of different structures. They are in principle analogous to standard optical microscopes, but instead of using visible light they exploit the wave properties of the electron. The advantage to using the electron is that its resolution is significantly higher. In a normal optical microscope where the range of wavelengths is from 380 nm to 750 nm the resolution limit will be around 200 nm. In comparison the human eye's resolution is around 0.1mm. For many semiconductor devices, where the sizes of some features are typically only a few nanometres thick, the resolution of an optical microscope is clearly impractical.

The de Broglie wavelength of an electron is of the order of several pm when accelerated in a potential of 100kV. Using this approximate wavelength and the resolution limit as $\frac{1.22\lambda}{D}$, where D is the aperture diameter, the TEM shows a much more suitable resolution of 0.2nm (500,000 times better than the human eye) when measuring nanostructures. Electrons are scattered due to their charged nature by the electron cloud and the nucleus of the atom. This is a fundamental difference between x-rays and electrons as a characterization technique (x-rays are only scattered by the electron cloud). By taking account of the scattering cross section of the electrons through the sample and the atomic scattering factor, the relative intensity of the diffracted beam can be determined. A more complete analysis of this is presented in section 3.3. The principles of both types of diffraction follow the same equations. Importantly the scattering intensity is dependent on this scattering factor and plane of diffraction.

Contrast is obtained through differences in core electron densities throughout the heterostructure. Materials such as Germanium are higher in density than Silicon so

will absorb/scatter more of the beam, appearing darker in the image. The cross section for elastic scattering is a function of Z (atomic number), which is why higher Z materials scatter more. This is known as Z contrasting. When imaging doped layers the change in density is generally very difficult to observe unless there is a suitably large number of dopants present. Therefore it is very unlikely to see contrast between a bulk layer and a subsequently doped layer.

The TEM uses an array of lenses for different stages of producing an image. The lenses used are electron lens, which are not dissimilar from those used in a conventional optical microscope in concept. Instead of glass they are made from an iron pole piece wrapped in wire. By applying a current, a magnetic field is produced which is used to condense or converge the beam. Unlike an optical microscope the lenses are not switched or moved during imaging. By varying the current through the lens the power may be altered to focus or magnify the image without physically moving the lens.

The TEM is represented in Figure 3.1. The whole column should be under high vacuum (below 10^{-7} mbar) in order to eliminate any electron scattering from gas molecules. At the top of the TEM column there is an electron gun. This is made from tungsten, and emits electrons through thermionic emission or field emission. To shield the filament and to focus the electron beam a Wehnelt cap is used. This is essentially a cylindrical cap with a small hole in the bottom. Below the cap is an anode to accelerate the electrons downwards. The cap is usually biased to a low negative potential to act to repel emitted electrons. This has the effect of condensing the primary beam slightly.

The beam then passes through the condenser lens to reduce the size of the beam before further passing through the condenser aperture. This aperture is fundamentally a small hole used to eliminate electrons from the beam. In this case it is used to collimate the beam, and reduce the width and intensity of the beam passing through. Next the beam will be incident on the sample.

Once the beam has passed through the sample it will next pass through the objective lens. This is used to expand the beam to form a magnified image, as well as focussing it on the phosphorus screen or CCD. This lens has a short focal length, and must have a very stable current supply, as this is what will determine the focal

length. An aperture may also be used here for blocking out high angle electrons and unwanted Bragg peaks when using diffractive mode. Other lenses below the objective lens and aperture are used to correct for aberrations of the beam as well as magnify the image. It should be noted that the effects of the collection of these lower lenses can be thought of as one intermediate lens providing several different functions. The final image is produced on a phosphorus coated screen, or alternatively a CCD camera, where images could be taken and recorded.

A key part of sample analysis using the TEM, which can prove invaluable, is the use of diffractive imaging. The general concept is to tilt the sample in the electron beam to select specific Bragg reflections which can provide further information about the structure. Under certain conditions strained layers or defects may be observable. The diffracted beam from the sample is deflected at specific angles relative to its crystal planes. The atomic scattering factor is related to these planes (described in section 3.3), which is proportional to the intensity ($I \propto F^2$). This means that only certain reflections are visible, and some reflections are more intense than other. Two useful reflections for TEM, used in this work, which have a large intensity are (004) and (220).

The image produced on the phosphorus screen is made up of all the possible Bragg reflections superimposed. Using the diffraction mode setting an array of spots, each representing a different Bragg reflection, is shown. As the sample is tilted using the goniometer electrons are diffracted from certain planes. By inserting an aperture to select the desired Bragg condition, along with the straight through [000] beam (which is formed by electrons passing through the sample un-scattered), a diffractive image can be observed. There are two different images in each different diffraction condition. The first is the bright field image, where only non-diffracted [000] electrons are selected, and the second is the dark field image, where only the diffracted [hkl] electrons are selected. In each case the electrons from the other spot are rejected, thus giving rise to contrast in each respective image. The reduced wavelength results in a larger radius of the ewald sphere (geometric construct used for electrons, neutrons, and x-rays). Therefore there are likely to be more Bragg peaks (resulted from crystal diffraction) intersecting the surface of the sphere, which is necessary for a Bragg peak to be observable.

Along with the pattern another feature can be seen in the diffractive mode, known as kikuchi lines. Secondary electrons may be scattered at the Bragg angle even when the beam does not satisfy the Bragg condition. Even though the number of these diffusely scattered electrons will be significantly lower than primary scattered electrons from the incident beam, there will always be secondary electrons which will diffract at the Bragg angle. As the electrons will enter from all directions a cone will be produced when observed in 3D. On the plane of the diffraction pattern a series of kikuchi lines will be observed where these cones intersect the Ewald sphere. kikuchi lines will be present from all potential Bragg reflections, therefore care and experience is required to identify the lines of interest.

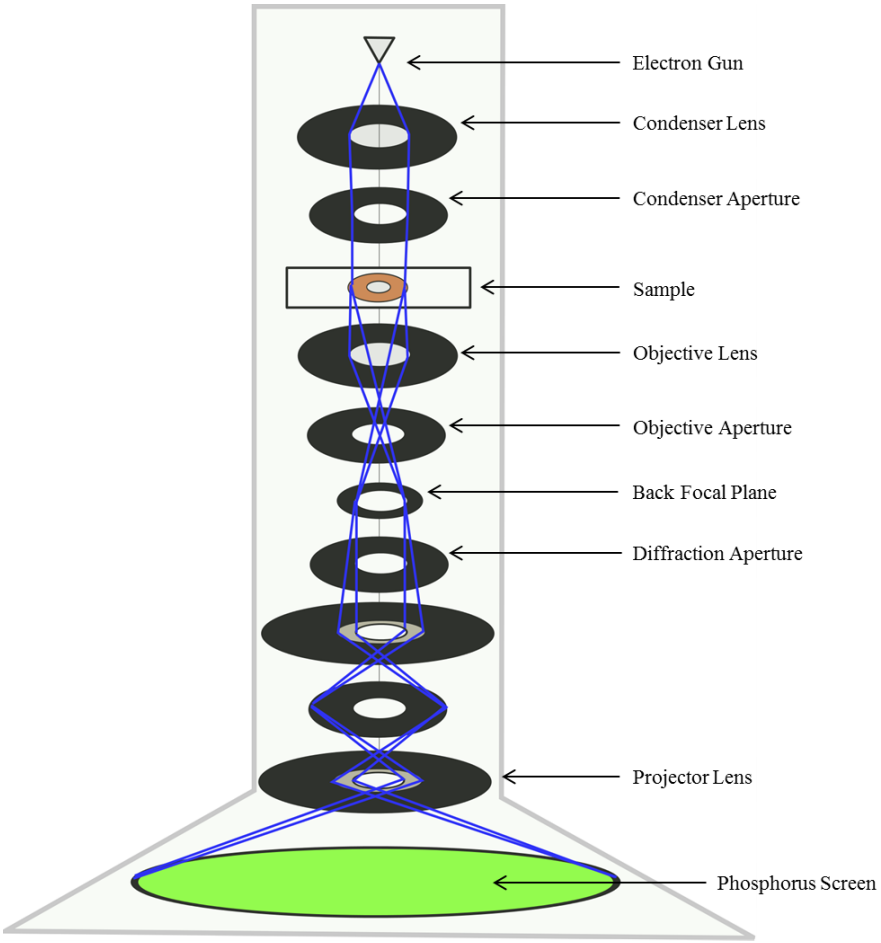


Figure 3.1: Schematic of TEM column. Electrons are generated at the top of the column before descending through the various lenses and the sample before producing an image on the phosphorus screen at the bottom.

An important concept for diffractive imaging is the burgers vector, which plays a significant role is imaging dislocations. As discussed in the previous chapter, dislocations arise from strain relaxation and their number should ideally be minimized. They are common at the interfaces of Silicon and Germanium which are present in SPADs, and have a detrimental effect on their performance. The burgers vector represents the displacement of the lattice. Under the condition $g.b = 0$ and $g.(b \times u) = 0$ (where g is the diffraction condition used, u is the line direction, and b is the burgers vector) dislocations meet the invisibility criterion. Strictly these products of vectors should be equal to zero, but actually dislocations will not be observed when these two products are actually less than or equal to 0.5. Misfit dislocations propagate along the [108] direction and the threading arms extend along the [109] direction, therefore the [220] diffraction condition is selected to observe these features. Also the invisibility criterion is generally not satisfied under this condition so is not appropriate to use. It should also be noted that, as stacking faults are also along the [109] direction, but made up from 30^o and 90^o partial dislocations which are invisible under the orthogonal diffraction condition, they may be distinguished from threading dislocations.

For the work in this thesis there were several different conditions used, namely straight through, (004), and (220). The two diffractions conditions used were done so to analyse the thickness of samples and observe dislocations (due to a high threading dislocation contrast) respectively. The dark field component observed the diffracted signal, while the bright field images showed the non-diffracted signal. Over the course of this work TEM imaging was required to assess: structural quality, presence of dislocations, layer thicknesses, and presence of highly doped layers.

3.1.1 TEM sample preparation

TEM sample preparation is a vital and often time consuming part in the process of imaging a structure. The samples must be very thin in order for a meaningful image to be collected. Typically they must be only a few hundred nanometres in thickness so that they are electron transparent. Ideally the majority of electrons which are incident on the sample should pass through, but unavoidably some electrons will be

scattered. Thicker samples will encounter more electrons scattering and an increase in multiple scattering events, which will detrimentally affect the image.

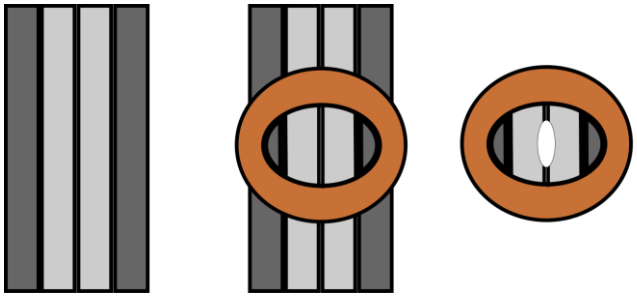


Figure 3.2: The three main stages of TEM prep. The grey and silver pieces are the glued gash wafer and the sample respectively. The copper ring is used to mount the sample into the TEM. Unnecessary wafer is removed prior to the milling process. A small hole is formed in the centre of the two pieces of wafer to obtain an electron transparent region.

Over one hundred samples were prepared and analysed during the course of this work so that final structures could be optimised wherever possible, or at least given a source for further work. Typical preparation time for a sample would be around 2/3 days. The basic preparation of a TEM sample involved; gluing two piece of wafer together along with two pieces of supportive wafer, mechanically grinding the samples down to produce as thinner layer as possible, and milling a thin hole in the sample using a precision ion polishing system. The area surrounding the hole in the sample will be electron transparent, and is the region which is observed in the TEM. Figure 3.2 shows the main steps of the TEM prep process used for this work.

3.2 Atomic force microscopy

Atomic force microscopy (AFM) is a surface technique used to map out the morphology of a sample. Despite relatively small scanning areas, typically of the order of $100\mu m^2$, this technique offers a very accurate measurement of surface roughness. This is of great importance for many structures as this may give an early indication as to the crystal quality of the growth, as well as suitability for further device fabrication. If a sample is particularly rough, techniques needed to form contacts to the surface often become very difficult to make and also reproduce. It is important to note that the scan taken from an AFM measurement is very small and only representative of the area observed. Therefore it is necessary to take a number

of scans from across the sample before any surface tendencies or features can really be described as representative of the whole sample. AFM measurements of samples in this work were averaged over 5 scans. A small sized scan may also result in significant deviation from the true nature of the surface. A useful analogy for this potential problem would be measuring a metre squared area on a mountain range to describe its overall roughness. Scans were also measured at an area of at least 10μm x 10μm, as the size of normal growth features should be significantly smaller.

AFM has a particularly high resolution in the z component where roughness measurements are able to measure down to only a few hundred angstroms. The key value obtained through an AFM scan is the root mean squared (RMS) roughness, but it also yields other important factors such as the minimum and maximum height across the surface. The model used for this work was a Digital Instruments Nanoscope III contact-mode AFM, with results analysed using Gwyddion software. A generic set up for the AFM which is used in this work is shown in Figure 3.3.

There are two universal modes for the AFM, namely contact mode and tapping mode. For both methods a small piece of sample (around 1cm^2) is positioning under the cantilever. On the end of the cantilever is a tip (most commonly made of SiN). The key elements to this technique are the; the tip and cantilever, the laser, the photodetector, and the piezoelectric stage. When using contact mode the tip is carefully brought into “soft contact” with the surface. The laser is incident on the cantilever and is reflected by use of several mirrors into the photodiode. As the sample is rastered under the tip the force on the tip will alter. When the tip is repelled or attracted towards the surface depending on the height of the feature it is traversing the cantilever will bend. This will deflect the laser spot on the photodiode signalling the surface variation from the set position. The piezoelectric stage will alter its position so that the spot is then shifted back towards its initial position and the force remains constant. The feedback loop will measure the voltage required which corresponds to a height change. Once the scan is complete a 2D and 3D image will be formed by combining the line scans.

Tapping mode uses the resonant frequency of the cantilever to measure the surface roughness. As the tip is rastered across the sample the cantilever oscillates at a

frequency close to the resonant frequency. When traversing areas which have a greater or lower height than that of the set point, the frequency will change. The feedback signal will alter the height of the piezoelectric stage to restore the frequency back to its initial state, recording the change in height. Tapping mode is often used for smoother samples with its higher resolution and minimal damage to the surface.

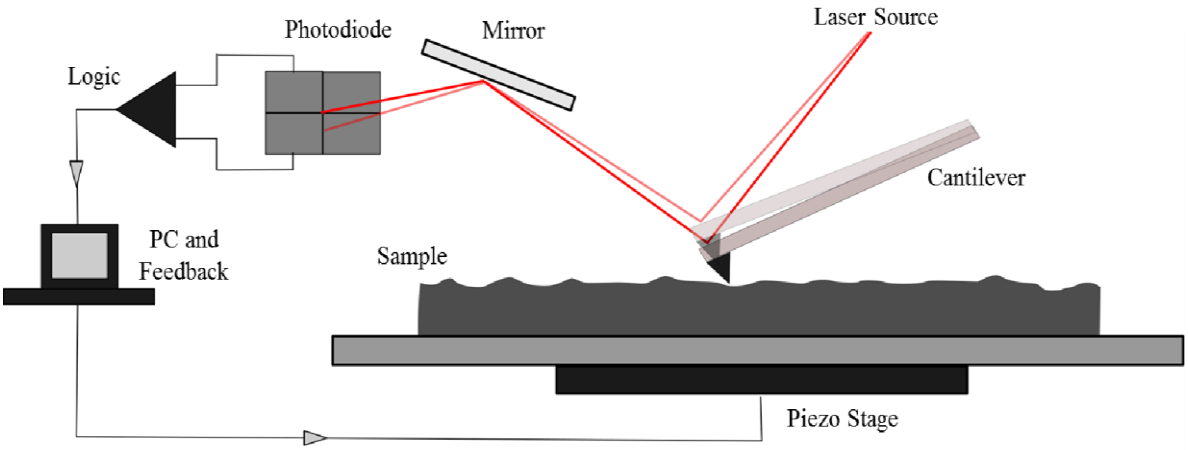


Figure 3.3: Diagram of basic set up for AFM. The cantilever and tip are rastered across the surface. The incident laser is reflected onto a photodiode, and the feedback system alters the position using the piezo stage, which is recorded as a change in height.

Focus on this work was in ensuring that the structures grown were suitably smooth for SPAD device fabrication. Generally tapping mode was employed to inspect a series of 10µm x 10µm areas for each sample. RMS roughness values were computed so that surface morphology could be assessed.

3.3 X-ray diffraction

High resolution x-ray diffraction (XRD) is an essential technique for measuring structural characteristics of a layer such as; lattice constants, strains, orientation, and defect densities. The x-ray analysis was performed, unless otherwise stated, using a Panalytical X'Pert PRO materials research diffractometer. It takes advantage of Bragg's law $n\lambda = 2dsin\theta$ to identify peaks corresponding to the presence of different materials in a layer. A copper source is used to produce a beam of x-rays, which can vary from 100 eV up to 1 MeV. The x-rays produced for

this work operated at 45 kV and 40 mA. The beam must then be monochromated using a series of Germanium crystals and collimating slits orientated to produce the desired wavelength. The beam is then incident on the sample where it will be diffracted. At specific angles Braggs law will be satisfied depending on the materials present in the sample. Another slit and monochromator are then used to collect only the diffracted beam. The set up of a standard lab based x-ray measurement system is shown in Figure 3.4 below.

There are five degrees of freedom in this x-ray set up. x, y, and z are used to position the sample in the beam so that the maximum intensity signal can be received by the detector, with a stage resolution of 0.1 mm. The other two degrees of freedom are the rotation of the sample, given by φ, and the tilt ψ, which both have a resolution of 0.1^{o}. The rotation and tilt of the sample are used to orientate the sample so that the required Bragg condition can be reached, and is shown in Figure 3.4.

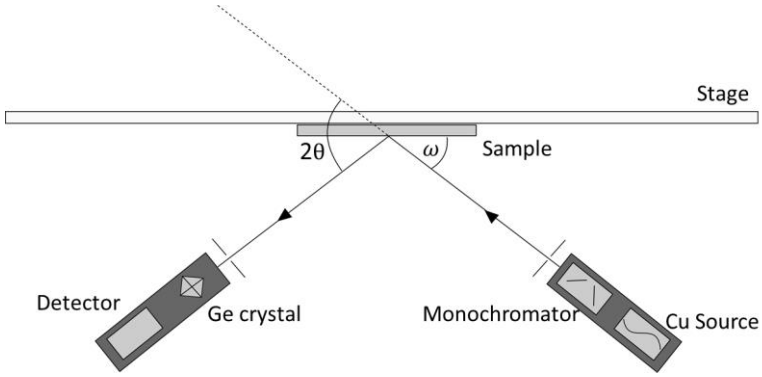


Figure 3.4: Typical x-ray kit set up. The sample is fixed to the stage, and the source and detector are moveable. The angles represented are ω (sample to source angle) and 2θ (sample to detector angle). A four bounce Ge crystal was used to produce a monochromatic source.

The two angles are omega (ω) and two-theta (2θ). Omega represents the angle of the incident beam with respect to the 001 plane of the sample. Two-theta is the angle between the incident beam and the diffracted beam.

A key to understanding the diffraction process is the concept of reciprocal space. To reach reciprocal space a Fourier transform is taken of the lattice space, which

gives rise to a reciprocal lattice. The reciprocal lattice can be described by a set of vectors, much in the same way as the crystal lattice can. For the crystal lattice the three vectors which define the structure are $a_{x(1)}, a_{y(2)}, a_{z(3)}$. For the cubic lattice in an unstrained state these vectors are all equal. The general reciprocal lattice vectors are described using these lattice vectors

$$b_1 = 2\pi \frac{a_2 \times a_3}{a_1 \cdot (a_2 \times a_3)} , b_2 = 2\pi \frac{a_3 \times a_1}{a_2 \cdot (a_3 \times a_1)} , b_3 = 2\pi \frac{a_1 \times a_2}{a_3 \cdot (a_1 \times a_2)}$$

Equation 3.1

When specifically looking at the diamond face centred cubic lattice seen in the group IV elements these equations can be simplified. In the unit cell all lattice vectors are perpendicular to each other. When taking a cross product of two unit vectors which are perpendicular the result is 1, therefore the reciprocal lattice vectors become

$$b_1 = \frac{2\pi}{a_1} , b_2 = \frac{2\pi}{a_2} , b_3 = \frac{2\pi}{a_3}$$

Equation 3.2

In the crystal lattice each point may be associated with an array of atoms. However in the reciprocal lattice these points represent the crystal planes. A useful construct in reciprocal space is the Ewald Sphere illustrated in Figure 3.5. The radius of this sphere is related to the wavelength of the incident x-ray beam by $2\pi/\lambda$. The origin of the reciprocal lattice is not the centre of the Ewald sphere, but instead lies on the circumference of it. An incident x-ray beam is represented by the radius of the sphere meeting the origin of reciprocal space. If a reciprocal lattice point lies on the surface/circumference of the sphere Braggs law is satisfied and diffraction can be observed. This represents a set of planes meeting the Bragg condition, and therefore a peak in x-ray data. The diffracted beam is described by the radius meeting the reciprocal lattice point, and the scattering vector Q connects the two k vectors.

Through trigonometry it can be seen that $sin\theta = (\frac{|Q|}{2})/(\frac{2\pi}{\lambda})$

As the equations for reciprocal space vectors show the relation to real space $|Q|$ may be instead written as $|Q| = \frac{2\pi}{d}$

By substituting these two equations it can be shown that the result is Braggs law.

The intensity of an x-ray signal is defined by its structure factor, $F(Q)$. Mathematically it may be written in the following form

$$F(Q) = \int_{UC} \rho_e(r) e^{(-iQr)} dr \qquad \text{Equation 3.3}$$

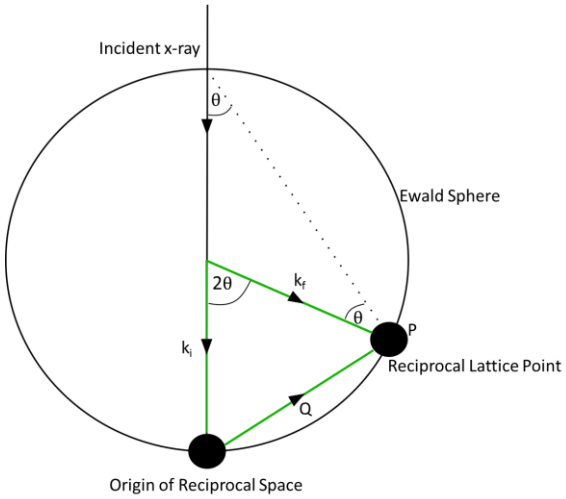


Figure 3.5: The Ewald sphere of reflection demonstrating the diffraction process. The radius of the sphere is 2π/λ.

This equation is integrating the charge distribution over a unit cell, where $\rho_e(r)$ is the charge density, Q is the scattering vector, and r the radius. It is easier to start by approximating this equation by instead looking at an single atom. The form factor can then be written in the following way

$$f = \int_{at} \rho_e(r) e^{(-iQr)} dr \qquad \text{Equation 3.4}$$

In this equation f is the atomic form factor. To find the form factor over the unit cell this equation simply has to be summed over *N*, where *N* represents the number of atoms.

$$F(Q) = \sum_{n=1}^{N} \int_{at} \rho_e(r) e^{(-iQ(r-r_n))} \mathrm{dr} \qquad \text{Equation 3.5}$$

The integral may be written by substituting the atomic form factor into this equation. $r_n$is the position of each atom in the unit cell. By replacing the atomic form factor section of this equation F simply becomes

$$F(Q) = \sum_{n=1}^{N} f_n e^{(-iQr_n)} dr$$ Equation 3.6

Recalling the laue condition which describes the diffraction of waves by the crystal lattice Q may be replaced, along with the full term for r_n

$$F_{hkl} = \sum_{n=1}^{N} f_n e^{(2\pi i(hx_n + ky_n + lz_n))} dr$$ Equation 3.7

From this point the specific type of lattice becomes important. For Silicon and Germanium the positions of the atoms in the face centred cubic lie at (0,0,0), $(\frac{1}{2}, 0, \frac{1}{2})$, $(\frac{1}{2}, \frac{1}{2}, 0)$ and $(0, \frac{1}{2}, \frac{1}{2})$. If these values are then substituted into the above equation for each atom in the cell, the structure can be written as

$$F_{hkl} = f_n\left(1 + e^{(i(h+l))} + e^{(i(h+k))} + e^{(i(k+l))}\right)$$ Equation 3.8

It is important to note that there are two main results from this equation depending on the reflection used. When each hkl are all even or all odd the resulting structure factor is 4*f*, and when they are mixed it gives a zero structure factor. Therefore this defines which reflections can give a Bragg peak.

As there are two other atoms in the diamond structure with coordinates (0,0,0), $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ a second criteria must be formed using the structure factor equation

$$F_{hkl} = f_n\left(1 + e^{\left(\frac{\pi}{2}i(h+k+l)\right)}\right)$$ Equation 3.9

In this case when the coordinates are substituted into the equation a non-zero structure factor is only obtained when the sum of hkl is equal to a multiple of 2n. Therefore, as $e^{ni\pi} = -1$ when n is odd, the second criteria enforces the sum of hkl to be an even multiple of 2.

These criteria allow for only certain reflections to be visible, and therefore it is important to select possible and intense reflections. The intensity of a given reflection is given by the square of the structure factor. The most commonly used reflections for Silicon and Germanium based materials which give valuable information are the (004) and (224). These are the symmetric and asymmetric scans respectively. The (004) reflection is used to analyse the out of plane components,

and the (224) reflection provides information about the in plane components. These reflections are chosen as they have a large structure factor for Silicon and Germanium, which is proportional to the intensity of the signal.

There are two main plots which can be obtained through XRD. The first is a rocking curve (RC) or an ω-2θ scan. A (004) rocking curve will provide information on the out-of-plane lattice parameter. It may also give a composition if the layers are strained. An ω-2θ scan is obtained by rocking the sample around omega.

As the sample is rocked the scattering vector q may intersect the Ewald sphere which gives rise to a peak, showing that Braggs law has been satisfied. The change in 2θ causes a change in the scattering vector. Strain fields associated with dislocations in a sample give rise to some unwanted scattering. A pure crystal would show a delta function peak, however this other scattering causes a smearing of the peak. Interference between the substrate and interface can cause some interference in the curve, which is seen as periodic fringes. The separation of these thickness fringes is proportional to $\frac{1}{\Delta d}$, where d is the thickness, and is used to quantify layer thicknesses.

A rocking curve demonstrates several important features. The two most intense peaks, for materials used in this thesis, correspond to the substrate (Silicon) and the Germanium layer. Any alloy layers will appear in between these two peaks. The larger the Germanium content the nearer the peak will be to that of Germanium. Any thickness fringes will be present on the corresponding peak. To calculate the thickness of a layer using the fringes seen on a rocking curve the following equation is used [78]

$$t = \frac{\lambda}{2\Delta\theta_p cos\theta_B} \qquad \text{Equation 3.10}$$

Where θ_p is the fringe separation and $\theta_B$is the Bragg angle for the desired reflection. This relation was essential for analysis of thicknesses of strained doped layers, necessary for the SPAD structure.

To see fringes the quality between the interfaces must be of high quality. X-ray rocking curves were measured using a PANalytical x-ray kit in this work, and fitted using a PANalytical X'Pert Epitaxy software package. If the layer becomes too thick and begins to relax the interface quality will be poor and therefore fringes will not be seen.

The other main type of scan utilised during XRD analysis is the reciprocal space map (RSM). For this set up the diffractometer should be in triple axis mode [79]. The initial beam is subject to a collimating slit and a Germanium Bartells 4-bounce (220) monochromator. The receiving optics is generally made up of a Xe proportional counter detector with a Germanium crystal positioned in front of it as an analyser crystal. The role of the Germanium crystal is to reduce the angular acceptance of the diffracted beam, which along with the collimator slit is able to describe a volume of reciprocal space. Therefore any Bragg peaks inside this volume will be visible on an RSM. An RSM is made up of many $\omega - 2\theta$ scans where ω is varied, which essentially means that an RSM is made up of multiple rocking curves.

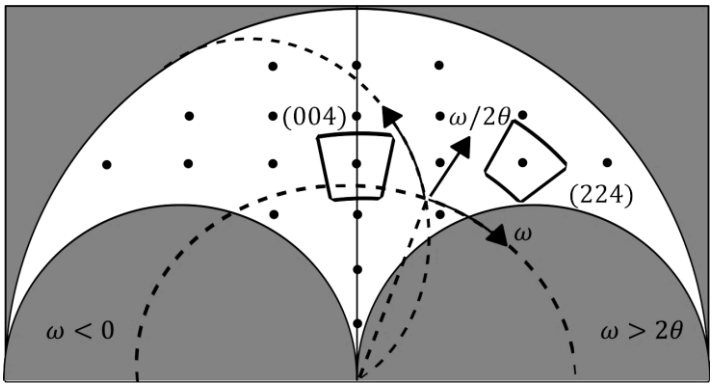


Figure 3.6: Schematic of reciprocal space, with potential visible Bragg peaks. Vertical axis represents [001] direction, and horizontal represents [108] direction. Adapted from [79].

The attainable area of reciprocal space is shown in Figure 3.6. The light area displays the Bragg peaks which are visible with the XRD kit set up according to Figure 3.4, while the darker areas cannot be obtained. The two small semicircles describe a situation where $\omega < 0$, and $\omega > 2\theta$, which upon inspection of Figure 3.4, it can be seen that these angles are physically impossible.

The two main angles used during XRD, ω and 2θ are represented by their respective dashed curves. These describe the arc mapped out in reciprocal space, and the resultant can be seen as the arrow labelled $\omega/2\theta$. As the ω value for each scan is altered an $\omega/2\theta$ scan is performed, which eventually results in a culmination of lots of scans mapping out a desired portion of reciprocal space.

Inevitably this method is time consuming in comparison to a single rocking curve. However there are many pieces of information that can be extracted from an RSM. These are namely: in- and out-of-plane lattice parameters using the (004) and (224) reflections, respectively, tilt, composition, and strain. While the lattice parameter may be calculated through only a (004) rocking curve, this is dependent on there being zero tilt in the strained layer.

The planes used for the majority of XRD rocking curves and RSMs are shown in Figure 3.7. As can be seen the (004) planes will provide the in-plane information, whereas the (224) planes can provide in- and out-of-plane parameters.

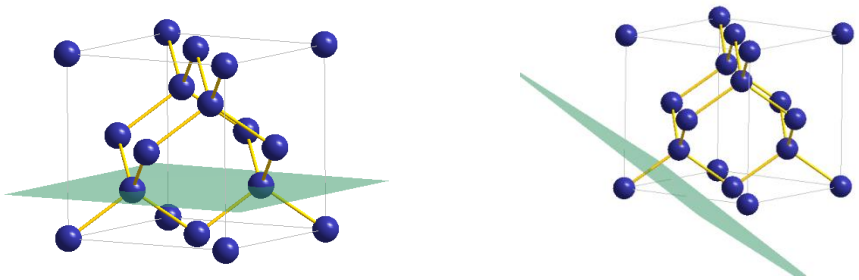


Figure 3.7: Diamond structure showing the (004) and (224) planes.

The position of the Bragg peaks is dependent on the material and composition of the layer. Materials with a larger lattice parameter than Silicon, such as Germanium will have Bragg peaks closer to the origin of reciprocal space. For alloys, the peak will lie between the two bulk peaks, depending on the exact composition. By obtaining a (224) scan the strain of the layer can be discovered. A fully relaxed layer peak will lie on the [224] vector, whereas the peak will lie to the left or right of this vector if the layer is under biaxial tensile or compressive strain respectively. This can be seen in Figure 3.7, where the area represented is that of the (004) and (224) areas from Figure 3.8.

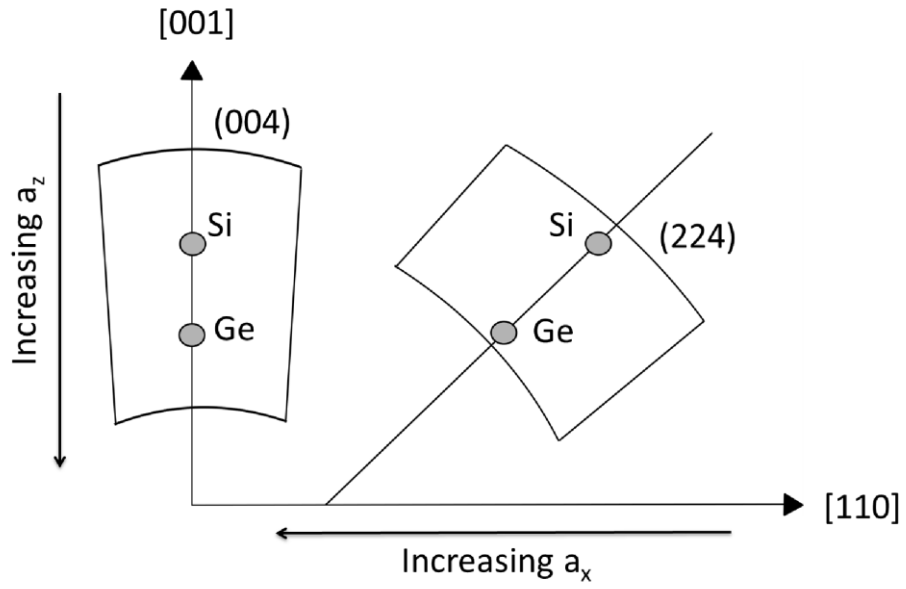


Figure 3.8: Diagram showing RSMs of (004) and (224) for Si and Ge. a_x and a_z are the in and out of plane lattice parameters respectively. Adapted from [79].

When a layer is strained to another, the Bragg peak will shift in the (224) reflection so that its corresponding a_x value is positioned in line with that of its preceding layer. The position of the Bragg peak can be used to determine the lattice constant of the layer.

3.3.1 X-ray measurements using synchrotron radiation

This section describes the technique of microfocus XRD for suspended SiB structures, detailed in Chapter 6, which were analysed at the Diamond Light Source, using Beamline B16 by the author.

For the standard lab based equipment used for bulk samples the beam size of around several hundred microns is more than sufficient. However this will be unable to detect any small features across the surface which could have potentially different properties. This is the case when attempting to investigate membranes which are often small in size. To explore these structures small x-ray beams are required, with spots focused to ~2.5 μm [80]. The main advantage of being able to use this small beam is that the sample area can be scanned to produce a map of strain or composition with micron resolution.

Suspended structures were measured for this work at the Diamond Light Source, which is a third generation synchrotron based in the UK. Beamlines here either use insertion devices or bending magnets, where the 3rd generation beamlines include insertion devices. However for B16, bending magnets are used. Although insertion devices are capable of very intense coherent x-rays, the intensity has the possibility of destroying delicate features such as membranes. Therefore bending magnet beamlines were more appropriate for this work. The principal of operation is bremsstrahlung radiation as the electrons in the ring accelerate around corners. During the cornering the electrons will undergo a strong deflection and therefore undergo acceleration. X-rays are produced at a tangent to this motion, in this case with 12.4 keV ($\lambda = 1$ Å). The maximum possible energy for the x-rays produced will be equal to the kinetic energy of the incident electron (eV).

The beamline optics are as follows; double crystal monochromator, toroidal mirror, double multilayer monochromator, and a set of Be compound refractive lenses (CRLs) [81]. These CRLs are used to focus the beam down to a size of the order a few microns, and are made up of a series of many concave lenses shown in Figure 3.9. The lenses are made from Be, which importantly has a low Z number, and therefore low X-ray attenuation coefficient, so that the loss in intensity of the beam over a series of many lens is small and not detrimental to the experiment.

It is important to keep a constant energy for the X-rays as the focal length of the lenses will change if it is altered. Unlike the lenses used for visible light, who have a refractive index much larger than 1, the refractive index of x-rays in most materials is around 1. This means that the focal length for such radiation is particularly long, which necessitates the need for many lenses. For the experiment at B16 the CRL set up was utilised to form a beam spot size of 3.72 μm x 1.85 μm. Previous experiments using CRLs at this particular beamline have been reported including that of Germanium membranes [82].

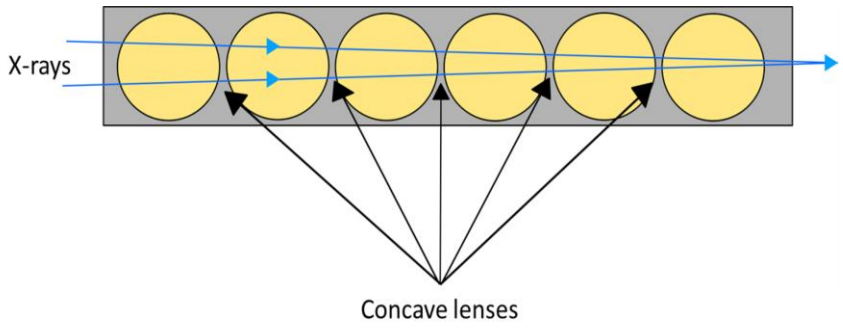


Figure 3.9: Schematic of compound refractive lens set up. Circles represent holes in the block. X-rays are focused to a point by the concave lenses.

SEM (scanning electron microscopes) stubs were used to attach the suspended structure before XRD analysis. This stub could then be easily mounted onto the measurement stage. This stage was capable of moving in each direction, X, Y, and Z, with an accuracy of 0.5 μm. This sample stage is mounted on the diffractometer. To try and reduce any signal loss from X-rays scattered in air a helium filled tube, as part of the detector arm, is positioned between the sample and the detector. The detector itself is a Pilatus 300 K area detector. This had an array of 487 x 619 pixels, where each pixel was 172×172 μm². Each pixel is a Silicon detector and has its own electronics for counting via the photoelectric effect. The Pilatus detector was prefered for this experiment for two main reasons. Firstly, the size of the detector is large, and was therefore able to take an RSM without having to physically move. Secondly the Pilatus detector will only measure events which exceed a set energy value, which removes noise from weak signals which would normally be detected.

For conventional X-ray kit 2θ must be scanned to find the optimum position. This is because the detector is a slit and is not capable of seing other angles. With the Pilatus detctor a wide range of 2θ values can be observed in one scan. The resolution of this is dependent on the size of the pixels on the detector and the detector-sample distance.

As with the lab-based XRD it was necessary to scan and optimise all of the diffractometer motors in order to maximise intensity.

3.3.2 (004) RSM with micro-focus diffraction

This section describes the procedure for obtaining a (004) RSM using the micro-diffraction process. Similar to conventional XRD alignment the sample is moved in the Z direction until the beam intensity is reduced to half, known as the half cut position. The angles and spatial axis were used in optimisation and measuring of the suspended structures.

The beam is then positioned using the X and Y components of the XYZ stage so that the X-ray beam is incident on the bulk material. This allows for a reference for the suspended areas to be compared to, by finding the (004) Si peak.

To locate the areas of suspended structures, such as the wires that will be described in Chapter 6, a series of coarse scans were performed. Initially the sample was scanned in x to find its location. From the intensity plot (tracking the Si peak) it was clear where the edges of the sample were as the intensity of the peak dropped to zero. The x coordinate was then set to the middle point of this x range. Next the sample was scanned in y to find the positions of the sets of wires. This was done by tracking the SiB peak. Areas which contained the wires would present a SiB peak, whereas the unprotected etched areas of bulk material would not see the peak. This allowed for a y coordinate for the wires to be established. The next step was to select the appropriate wires. With the general coordinates for a row of wires established the samples was scanned along x for the SiB peak with a finer step size. Over the wire a SiB peak would be present, but as the scan passed over a gap the peak would disappear. This allowed the appropriate sized wires to be found for analysis, along with checking that they were structurally intact.

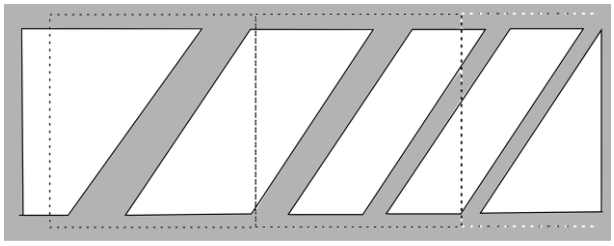


Figure 3.10: Schematic of 100μm x 100μm areas mapped out across set of suspended wires. The wires (grey bridges) were fabricated with varying thicknesses.

For the experimental scan the XYZ stage was used to position the suspended structure in the necessary area. This stage featured a 25 mm range with 500 nm precision. A piezo stage was mounted on top of this stage, which had a 100μm range and 50nm precision. This piezo stage could then be used to map out a 100μm x 100 μm region before the X, Y, Z stage could move to the next position. Figure 3.10 shows the dashed areas which could be moved to firstly with this stage, while and then could be more accurately mapped with the piezo stage. Figure 3.11 shows a cross section of this set up. During a piezo scan positions over the square were mapped every 5μm.

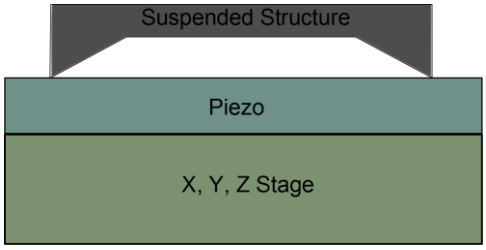


Figure 3.11: Schematic showing the X, Y, Z stage with piezo stage and sample. The suspended sample is positioned on top on the piezo stage and can be moved precisely.

Instead of varying ω at each point on the 100 μm square map, each point is imaged with one value of ω before it is incremented by 0.01^{o}. This is preferred as the piezo map is much faster than the conventional technique. The range of ω is chosen so that it covers both the Si and SiB peaks. The result of this process is that each spatial point will have an image for every ω value in the range.

Analysis of the data was performed by taking the tif image from each directory corresponding to one spatial position on the 100 μm map. With each tif image from each ω value for a certain spatial position, a full RSM over the ω range was completed. Once a full 100μm region was scanned the coarse stage would be used to move to the next 100μm space, therefore piecing together a map of all the wires. With RSMs taken at every spatial point a map of RSMs can be obtained for any given 100 μm region.

The equations used to convert from the physical angles to reciprocal space were

$$q_x = \frac{4\pi}{\lambda}\sin(\theta)\sin(\omega - \theta)$$ and Equation 3.11

$$q_z = \frac{4\pi}{\lambda}\sin(\theta)\cos(\omega - \theta)$$ Equation 3.12

Localised, real space changes in strain and tilt can then be obtained from these RSMs. For each spatial position, information on strain and tilt can be found by summing the RSMs over q_x and q_z respectively, thereby reducing the RSMs to diffraction profiles as function of q_z and q_x. From these equations information such as the strain as a function of position can be obtained. The q_x must be summed over the q_z direction, and the q_z must be summed over the q_x direction to produce an intensity plot for each q_x and q_z. With this information a gaussian can be fitted to the curves to find the peak positions for each vector. The summing process essentially compresses the data into a 2D form where the intensity across each value dictates the height at each point on the plot. This 2D plot is easier and faster to fit, as opposed to the 3D set.

Because the Si peak will be significantly more intense than the SiB peak (due to a larger diffraction volume) the perpendicular component must be fitted with two separate curves. Firstly a curve will be fitted to the Si peak, and then the same process will be used to identify the peak in that of the SiB.

With this process a (004) scan using micro diffraction at B-16 can be used to identify the out-of-plane component and the tilt for each position along a suspended wire.

3.4 Hall effect measurements

Hall effect measurements were carried out and fabricated in the Nano-Silicon clean room by the author using a custom built cryostat system. Calibration of this kit was performed prior to the measurements using samples with known doping concentrations. The basis of electrical measurements performed in this work is the hall effect. A carrier moving through a piece of semiconductor material with an applied field perpendicular to its motion experiences a force which is normal to both the motion and field. This is called the Lorentz force and is depicted in Figure 3.14. It is given by the equation

$$F = q(E + v \times B)$$ Equation 3.13

In the Figure 3.12, the current is flowing through the bar shaped sample from left to right. The applied magnetic field is into the page, which means that the Lorentz force is upwards on the bar. As these electrons flow across the bar the force pushes them upwards causing a drift of electrons to the top of the bar. As the electrons begin to build up along the side of the bar a negative charge becomes apparent there. Also along the bottom of the bar there becomes a lack of electrons, which results in a build-up of negative charge. The overall result of the two charged sides of the bar is a potential difference. This is the Hall voltage and its magnitude is given by

$$V_H = \frac{IB}{qnd}$$ Equation 3.14

where *I* is the current, *B* is the field, *q* is the charge of the electron, *n* is the carrier density, and *d* is the thickness.

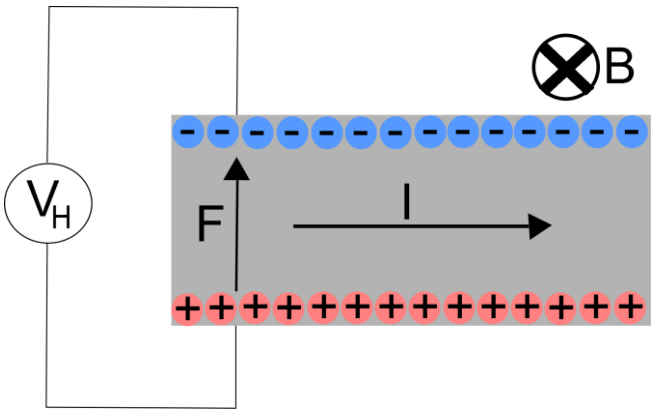


Figure 3.12: Schematic of a Hall Effect measurement.

This quantity is incredibly useful as it can be used to calculate other key parameters. To calculate the sheet density the following equation is used

$$n_s = \frac{IB}{q|V_H|}$$ Equation 3.15

The carrier density can also be calculated if the thickness of the layer is known (from TEM or another technique).

The sheet resistance and mobility of a sample may also be of interest and can simply be calculated using

$$\mu = \frac{V_H}{R_s I B} = \frac{1}{q n_s R_s} \qquad \text{Equation 3.16}$$

These results can be obtained using a Van der Pauw set up [83]. This is a popular technique for electrical measurements of semiconductor materials due to their convenience. The basic set up which was used during this work is as follows. It should be noted that the shape used may vary. A VdP measurement must adhere to a few practical parameters. The contacts must be ohmic (before each measurement IVs are performed to check), the sample must be uniform and without any holes, and the sample should be measured in dark conditions to avoid any photovoltaic effects.

A square piece of sample was cleaved and aluminium contacts were evaporated onto the surface using a mask. This gave a series of squares marked by dots of metal. One square was cleaved from this with metal contacts at each corner. The sample was then loaded into a chip package ready for measuring in a closed cycle cryostat (CCC) system. To obtain the desired characteristics of the sample a resistivity and Hall measurement had to be performed. A resistivity measurement is used to obtain a sheet resistance, and with that a hall measurement could be used to find the carrier concentration. To find the sheet resistance a DC current was applied between contacts 1 and 2 (opposite corners of the square), before measuring a voltage between contact 3 and 4 (other corners of the square). The same is repeated between contacts 2 and 3, and, 1 and 4. Therefore the sheet resistance may be calculated using

$$R_A = \frac{V_{12}}{I_{34}}, R_B = \frac{V_{23}}{I_{14}}, e^{\frac{-\pi R_A}{R_S}} + e^{\frac{-\pi R_B}{R_S}} = 1 \qquad \text{Equation 3.17}$$

The convention used for measuring the hall voltage is to force the current between contacts 1 and 3, and therefore measure the voltage between contacts 2 and 4. Once the hall voltage has been acquired it is then possible to calculate the sheet density.

It is important to apply an AC current when performing such measurements, as this removes the Nernst effect. When applying a DC current hot electrons will drift to

one side of the sample while cold electrons move to the other. This introduces a small unwanted voltage independent of any applied field.

3.5 Growth techniques

The most common way in which semiconductor devices are produced is via epitaxy. Layers of the required materials can be reproducibly grown to strict specifications before they are subject to further fabrication. There are two widely popular examples of systems used for epitaxial growth; molecular beam epitaxy (MBE) and chemical vapour deposition (CVD). During MBE growth highly pure elements are heated until sublimation occurs, allowing the elements to be transported to the surface of the wafer where they condense, forming single crystal layers. In CVD growth gaseous compounds containing the required elements pass into a heated chamber, depositing layers of material as the molecules dissociate at the wafer surface. CVD growth has the advantage of mass production and is generally used as an industry tool. Vast numbers of samples can be produced in a short space of time, allowing for a quick turnaround. Samples were grown for this thesis using a reduced pressure chemical vapour deposition (RPCVD) system. Detailed analysis of each sample was required to optimise the growth of future samples.

3.5.1 Precursor gases and growth modes

Epitaxial growth of layers using CVD requires large quantities of different gases. These gases contain the elements required for given layer. They are supplied as a mixture of at least one other gas, known as the carrier gas. Probably the most common carrier gas is hydrogen, although this is certainly not the only option. Silicon must bond to surface atoms to be incorporated into the lattice, which means that hydrogen must be desorbed from the surface. It can therefore be useful to vary the temperature during growth (and therefore rate of hydrogen desorption) to increase or decrease the growth rate of the Silicon. To increase the growth rate even further an alternative carrier gas such as nitrogen, which is not a surfactant, could be used instead of hydrogen. When a precursor gas enters the growth chamber it will be broken down into other compounds appropriate for the growth. The most common precursors for Silicon and Germanium are silane (SiH_4) and germane (GeH_4), respectively. The basic reaction which takes place at the surface

involves producing an adatom of either Si or Ge which is mobile and can be incorporated into a lattice site. This occurs via the process:

$$SiH_{4(g)} \leftrightarrow Si_{epi} + 2H_{2(g)}$$

$$GeH_{4(g)} \leftrightarrow Ge_{epi} + 2H_{2(g)}$$

where heat is required to produce either phase. Germanium growth may occur at as low as 280^{o}C [22].

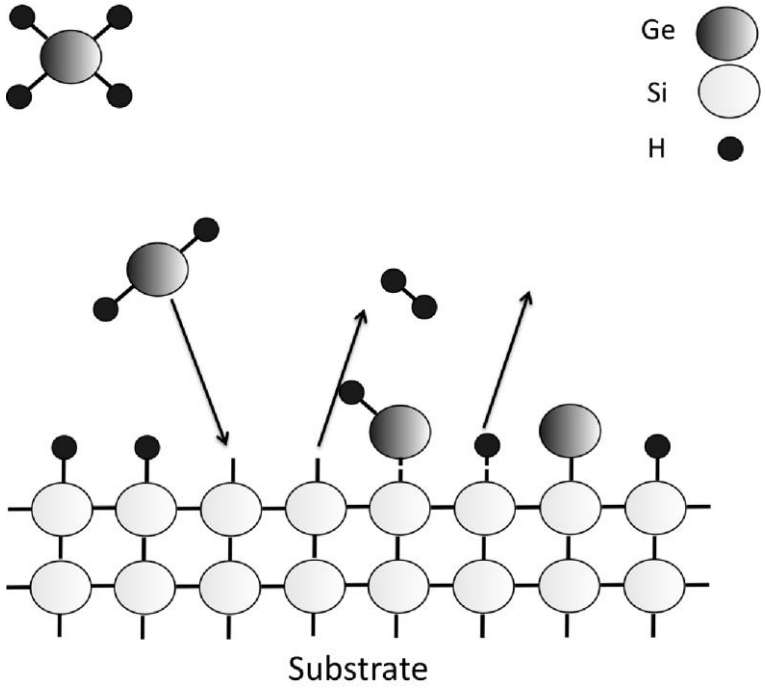


Figure 3.13: Diagrammatic representation of precursor gas reacting with the surface to form epilayers. The gas molecules are incident on the surface before they are absorbed and any reactions products are removed from the chamber. The dashed bonds represent adatoms which have not yet been absorbed into the lattice.

Alternative precursors for Silicon and Germanium growth are disilane (Si_2H_6), dichlorosilane (SiH_2Cl_2) and digermane (Ge_2H_6). Importantly these precursors are useful for growing at lower temperatures than is possible with SiH_4 and GeH_4. Importantly, these precursors such as disilane are more reactive than silane and can be used to grow out of equilibrium. This means that the adatoms do not have enough time to move far before the succeeding layer is absorbed. Growth rates are

much faster than silane which allows for preceding layers to be “locked in”. However for these precursors the increased reactivity can lead to the bonding of atoms in the gas phase which can result in the promotion of 3D growth. For the commonly used dopants, boron and phosphorus, the gases diborane (B_2H_6) and phosphene (PH_3) [23] are used.

Figure 3.13 represents the basic reaction which takes place between the precursor gas and the surface on which growth is to take place. Typical growth stages are described in Figure 3.14 to demonstrate how the precursors are brought into the CVD chamber and result in epitaxial growth.

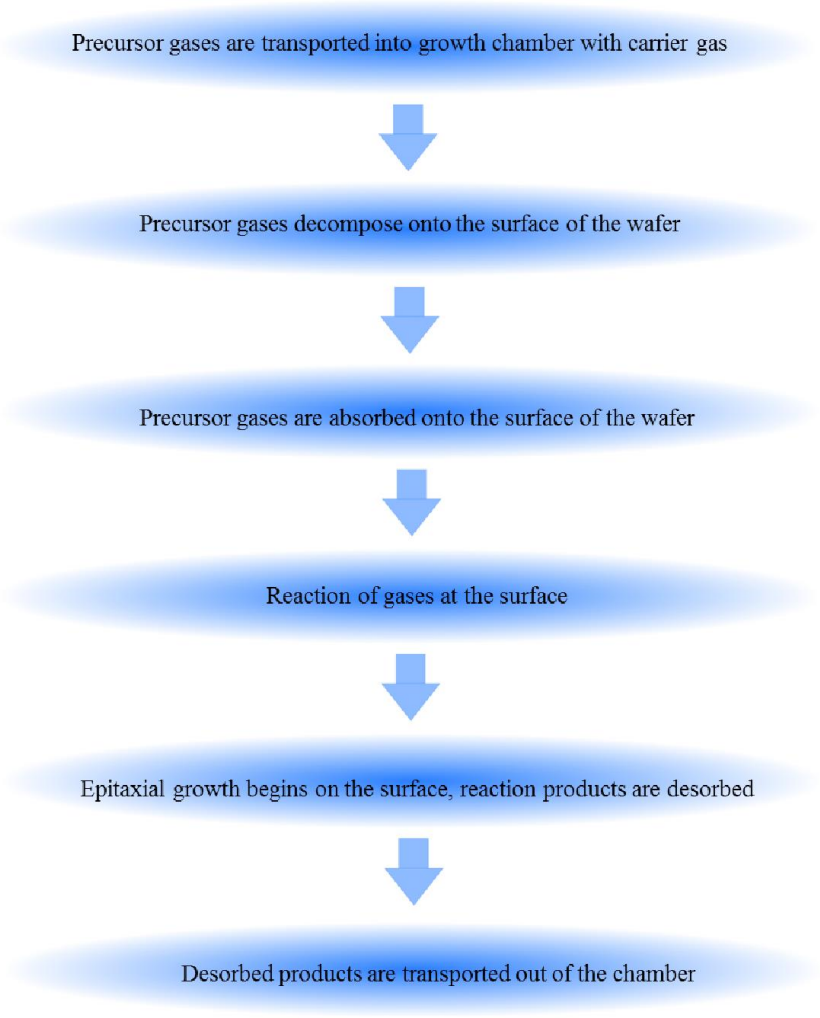


Figure 3.14: Six major steps for deposition using RP-CVD.

There are three different growth modes by which epitaxial layers are grown, and is shown in Figure 3.15. These are namely, 2-D Frank van der Merwe [24], 3-D Volmer Weber [25], and Stranski Krastanov [26]. Ideally an adatom will be incorporated at a surface step site. For a homoepitaxial layer there is no mismatch between that of the layer and the one preceding it so there will be no strain induced. This means that the layer is able to grow via the Frank van der Merwe mode whereby the growth is layer by layer shown below.

If there is a large amount of strain in a layer (due to heteroepitaxial growth where there is a large mismatch), or the thermal energy is low (the adatoms are not very mobile) the adatom may be incorporated alone on the surface. This can lead to other adatoms attaching themselves to this atom and create a cluster of atoms on the surface, as this is energetically favourable. This 3-D island growth is termed Volmer Weber mode. It is possible for the growth to start out as Frank van der Merwe when the amount of strain is low, but subsequently develop into Volmer Weber when the accumulated strain exceeds the limit for 2-D growth. Hydrogen is used as a surfactant during the growth period. It acts by reducing the surface free energy which in turn leads to improved 2-D growth.

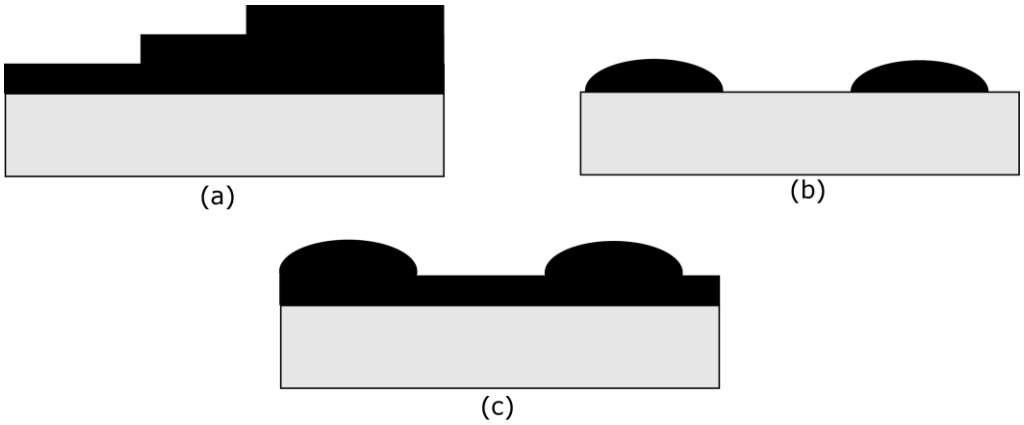


Figure 3.15: The three different growth modes, Frank Van der Merwe (a), Volmer Weber (b), and Stranski Krastanov (c) respectively.

The mode of growth can be explained using the free energies of the constituent regions. 2D growth will be observed when adatoms preferentially attach to the substrate which has a greater free energy, or $\sigma_s > \sigma_i + \sigma_l$, where σ_s, σ_i, σ_l, represent the free energies of the substrate, interface, and layer respectively.

However, when the free energy of the substrate is lower than that of the interface and substrate $\sigma_s < \sigma_i + \sigma_l$ the adatoms will favour attaching to other adatoms rather than the substrate and create mounds/3D growth.

There are two growth regimes for chemical vapour deposition. These are namely temperature limited and mass flow limited. The temperature limited growth regime describes a low temperature situation where the growth rate is limited by how fast the (temperature dependent) deposition reactions take place. At higher temperatures the limiting factor switches from the reaction energy to how fast the precursor gases can be supplied to the substrate – this is the mass flow limited regime.

3.5.2 Growth kinetics

When atoms are deposited onto a substrate they are known as adatoms. These adatoms are able to traverse the surface until they are chemiabsorbed at a surface site or defect. The temperature at the substrate defines how much energy the adatom has to move. This, along with several other factors gives the adatom its migration length (generally μm), which defines how far it may travel. At higher temperatures the adatoms have more energy to overcome the energy barriers present between surface sites. Once the adatom no longer has the energy to overcome the barriers it becomes incorporated as an atom.

Substrates used in industry and for structures grown during this work have atomic terraces present. These terraces have dangling bonds which promote the absorption of the adatoms and therefore increase the likelihood on 2D growth. To maintain two dimensional growth the migration length should be greater than that of the terrace width. If the energy of these adatoms is too low they will be incorporated at a surface site, which can then become a site for further 3D growth.

3.5.3 Substrate preparation

Prior to epitaxial growth it is essential that the substrates are in excellent condition: a smooth and clean substrate is vital to obtain perfect single crystal epitaxial layers. Firstly, the wafer must be smooth. For all the material used in this study, the root mean square (RMS) surface roughness of the wafers was around 0.1nm. Secondly, to ensure a clean surface for growth any contaminants or oxides which could be

present on the substrate had to be removed. If any metal or organic contaminants needed to be removed wet chemical etching would be performed. However, the majority of wafers were never removed from a clean room environment prior to or during the growth processes so the only substrate preparation necessary was the removal of the native oxide. Desorption was performed by using an in-situ bake of the substrate at 1150^{o}C, under hydrogen gas flow to ensure that the surface was passivated with hydrogen. With the wafer in this state growth could then be initiated without risk of contaminants ruining the growth quality.

3.5.4 Reduced Pressure CVD

CVD systems are used for many different applications, which rather expectedly require a vast array of different conditions for growth. Two of the primary variables are growth temperature and growth pressure. Growth can take place over a wide range of temperatures (100^{o}C to over 1700^{o}C) and pressures (ntorr to atmospheric (1013mbar)) depending on the material of interest. The types of CVD which are in use today include UHV-CVD (ultra-high vacuum), AP-CVD (atmospheric pressure), and RP-CVD (reduced pressure). The RP-CVD used to grow the samples in this work is an ASM Epsilon 2000 reduced pressure CVD.

Figure 3.17 depicts the cross section of a typical RP-CVD system. Precursor gas flows into the chamber via the pipe line on the left. Valves are used to select which gas is flowed into the chamber, as well as the gas flow rates. The growth of the epilayers is performed inside the growth chamber. This is made using a suitably transparent material, often quartz, so that that the chamber can be irradiated by a series of infra-red lamps which surround the outside walls. Temperature is accurately controlled using thermocouples underneath the wafer itself to allow for precise control of the growth. These are used to control the power output of the lamps and can be adjusted according to specific growth plans. The Epsilon reactor used during this work is a cold wall reactor. This has the benefit of reduced temperature on the walls which should reduce unwanted deposition, but does result in a slight temperature variation across the wafer itself.

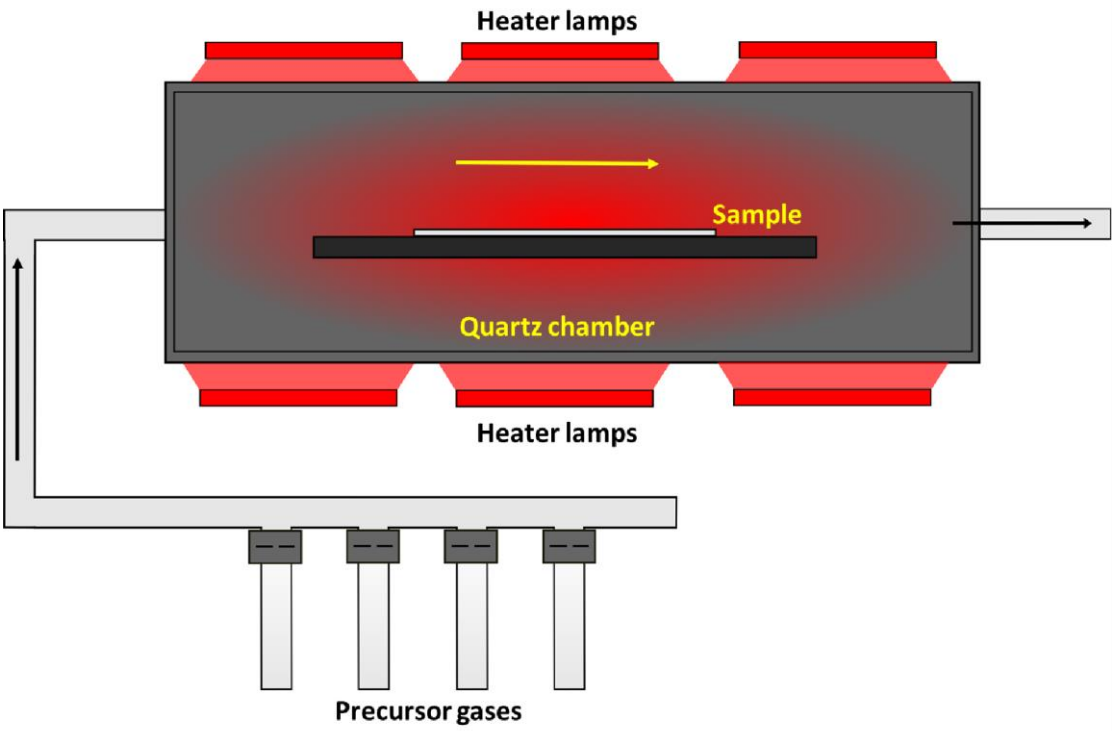


Figure 3.17: Schematic of an RP-CVD system. The lamps are used to heat the chamber and substrate. Precursors are passed through the chamber where the reactions for growth take place, before the waste gas is removed through an exhaust system.

3.6 Secondary ion mass spectrometry

Secondary ion mass spectrometry (SIMS) is a destructive characterization technique used to obtain a depth profile of a structure. The main results which are gained from SIMS analysis are doping concentrations, layer thicknesses, intermixing of layers, and diffusion tails. High (low) energy ions between 1 and 30 keV (500 meV) sputter the surface of a very localised region of the sample (typically around 250 μm^2), depicted in Figure 3.18, which produces positive, negative, and neutral ions. It is important to note that the energy must be sufficiently high to ionize the surface of the structure. Secondary ions, which have been sputtered from the surface, are then subject to a mass analyser which will identify which elements are present in the sample. A typical method for analysing the sputtered ions is through a quadrupole analyser. Other techniques for SIMS analysis utilise either time of flight or the variation in curvature of the ions path in a magnetic field. The quadrupole analyser technique is comprised of 4 poles, 2

opposing rods with a positive charge, and tow with a negative charge. The (charged) sputtered ions are oscillated by the field set up through these poles the mass to charge ratio can be tailored so that only ions in a given range may be detected. The ions outside of this range are rejected through collisions with the poles as they oscillate between them. For a sample to be analysed through the SIMS technique it is important to know what atoms should be present in the layer so that the equipment can be "tuned" accordingly.

Depth profiling is obtained by knowing the sputter rate of the beam and measuring the intensity of a given mass signal as a function of time [84]. The sputter rate is calculated by measuring the depth of the crater formed along with the time taken. This will show the concentration of the different elements present at specific depths. The resolution of this profiling is dependent on the uniformity of the beam etch, and the mass and energy of the ions. High energy SIMS which is the most commonly used technique throughout this work generally has a faster sputter rate to that of low energy SIMS. Despite an improved resolution low energy SIMS is a more expensive technique due to the slowness of sputtering so has only been used where appropriate.

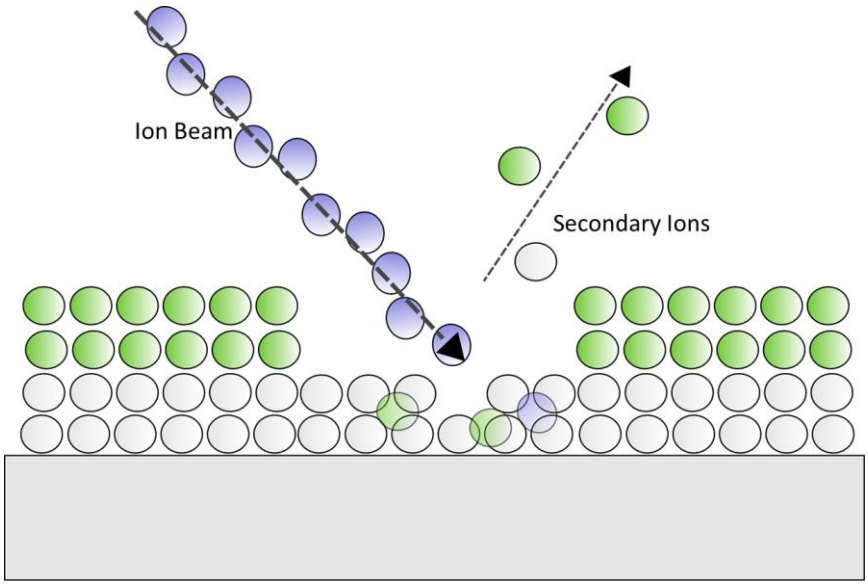


Figure 3.18: SIMS, the surface is sputtered by a primary ion beam, resulting in ejection of secondary ions.

The ions used for the sputtering process in current practice are noble gases, most commonly argon, oxygen, and caesium. Important atoms used in this work for doping are boron, phosphorus, and arsenic. During SIMS analysis boron produces a positive ion, whereas phosphorus and arsenic produce a negative ion. Caesium lowers the work function which allows for easier ejection of electrons. This allows for more group V (i.e. P and As) atoms to be filled, therefore producing negative ions. To measure boron concentrations oxygen ions are the most commonly used. They have a high electron affinity which is characterized by the "grabbing" of electrons from the boron atoms, resulting in a positive boron ion.

There are a few factors which limit the reliability of a SIMS profile. Two important factors which are relevant to heterostructures are the surface roughness and the density of threading dislocations. Rough surfaces may deflect secondary ions and reduce the number which are measured by the mass analyser. This limits the accuracy of absolute values obtained. Threading dislocations may cause faster sputtering in regions with a large population, resulting in an uneven beam sputter. This limits the accuracy of depth measurements. These factors are dependent on the quality of the growth of structures so can be reduced if possible. One factor which is a characteristic of the technique itself is where the ion beam pushes lattice atoms further into the sample [85]. When this occurs these atoms may produce secondary ions at a later time which does not represent their original location.

SIMS profiles for this work were measured at EAG labs. Doping profiles were obtained using the ions described above. Data analysis was performed in house to obtain values such as doping concentrations and diffusion tails.

Chapter 4

Development of Ge on Si SPAD devices

4.1 Single photon avalanche diodes

This chapter explores the designs of Silicon and Germanium SPAD structures. Optimization of each layer is necessary for minimising potential dark counts and ensuring the maximum chance of efficient absorption and transport of carriers for multiplication. Doping is given particular attention as abrupt and accurate doping profiles are essentially for SPAD operation. Electrical measurements are performed to explore the dark count characteristics which often plague Germanium layers in SPADs. Various different designs are explored and a preferred final design has been selected following discussion with collaborators.

The structures proposed in this work require separate absorption and multiplication SPADs where Silicon is used as the multiplier, and Germanium is used as the absorber. The concept of this design is that a photon is absorbed in a region specifically chosen for the desired photon wavelength, and then undergoes impact ionization in a separate region which has advantageous avalanche characteristics. Most follow the same design as shown in Figure 2.22, but will be described in further detail where appropriate in the upcoming sections. The doping profiles throughout the structures should ensure that the there is a high electric field in the Silicon region, so that avalanche breakdown can occur, and that there is a low electric field in the Germanium layers, to ensure drift of carriers into the multiplication region. CVD was used to epitaxially grow the layers required before any fabrication. Aspects of the growth had to be optimized to allow effective operation of the SPADs. This section explores the issues involved with acquiring perfectly grown samples, and some electrical testing of the device made in light and dark conditions.

4.2 Optimization of absorption and multiplication regions

As previously mentioned there are two distinct regions of intrinsic material required for a SPAD design. These are the absorber and multiplication regions, which for this work used Germanium and Silicon respectively. While the growth of these may be assumed to be straight forward, the optimization of such layers is far from simple. Several growth parameters such as defect density, doping concentrations, and diffusion of atoms can be affected by the growth method. The

following sections describe the growth of the intrinsic layers used for the various group IV SPADs explored in this work.

4.2.1 Growth of Silicon multiplication region

Separate absorption and multiplication SPADs require perfect crystalline absorber and multiplication regions to reduce transit times for carriers as well as minimise possible sites for trapping and dark counts. Each individual layer had to be investigated in order to ensure appropriate growth conditions were applied when growing the structures as a whole. The homoepitaxy of Silicon growth on a Silicon layer (or substrate) should not lead to any dislocations forming through relaxation, due to the obvious reason of equal lattice constants. Despite the lack of issues experienced during Silicon homoepitaxy, layers were grown and analysed for calibration purposes. Growing the layers too quickly could lead to polycrystalline growth ruining the final device. However, the growth should not be so slow that the growth of these thick layers restricted the potential for production at an industrially acceptable rate. To demonstrate the potential problems with a non-clean substrate Figure 4.1 is presented. Particulates on the surface may lead to nucleation sites for polycrystalline growth, and clearly should be avoided.

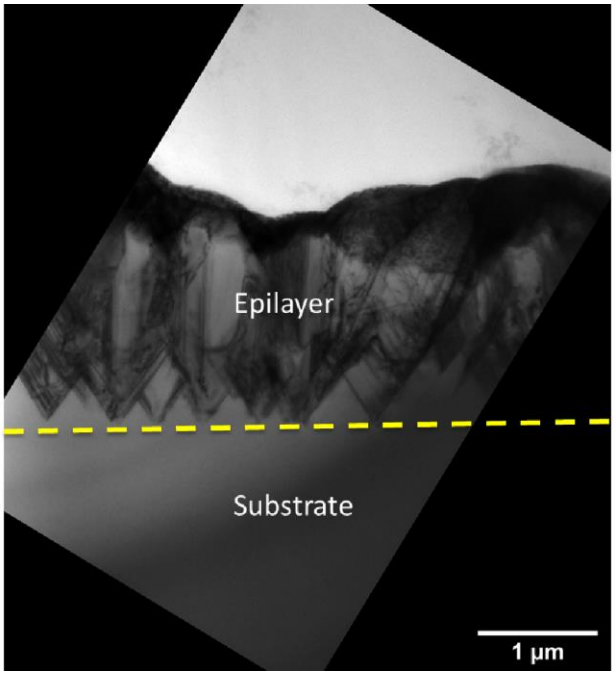


Figure 4.1: Cross-sectional (220) TEM image in dark field conditions demonstrating the dark areas where growth is becoming non crystalline.

To ensure that the Silicon growth was suitable for the multiplication region several calibration samples were grown. These structures comprised of Silicon layers of varying thickness (growth times) with thin Silicon-Germanium alloy layers used to mark the end of each period. Thicknesses of these layers were kept thin enough to avoid any dislocations. A spread of different growth temperatures from 500^{o}C up to 800^{o}C was used so that further growth rates could be predicted based upon these results. Figure 4.2 shows an example of the structures used to calibrate the multiplication region growth.

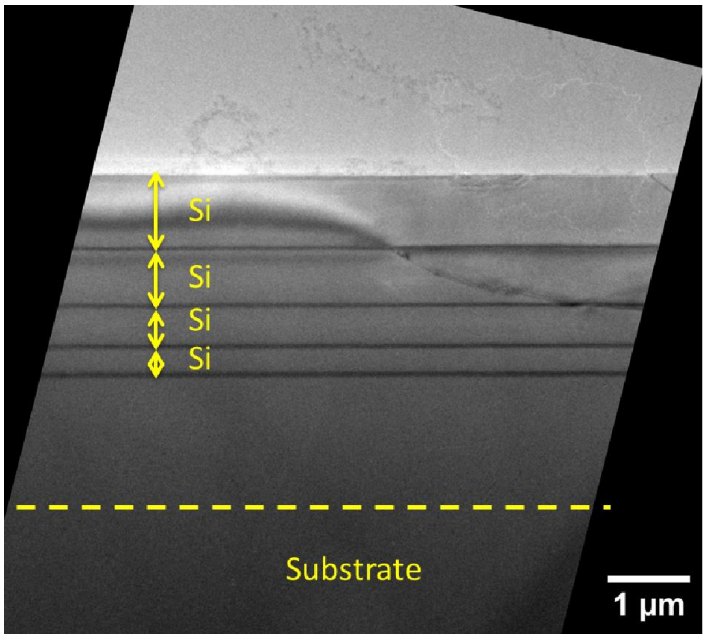


Figure 4.2: Cross-sectional TEM image of Silicon/Silicon-Germanium multilayers grown at a temperature of 700^{o}C. The thin dark lines are SiGe spacer layers. Growth of the Silicon was optimized at temperatures ranging from 500^{o}C up to 800^{o}C.

An important consideration is the uniformity of the epitaxial growth. The wafers are positioned in the CVD chamber so that the growth conditions should be as constant as possible across them. However, inevitably there are variations in thickness across a wafer. Applying this logic to a full structure grown for device fabrication, some areas of a sample may be significantly different. This is an issue for many semiconductor devices that rely on specific thicknesses for operation. To highlight this issue Figure 4.3 shows a line scan of the surface of a Silicon layer grown on a doped Silicon substrate. This measurement measured the Silicon

epilayer, rather than simply the height of the surface, as this could be misinterpreted if there was any bowing of the wafer.

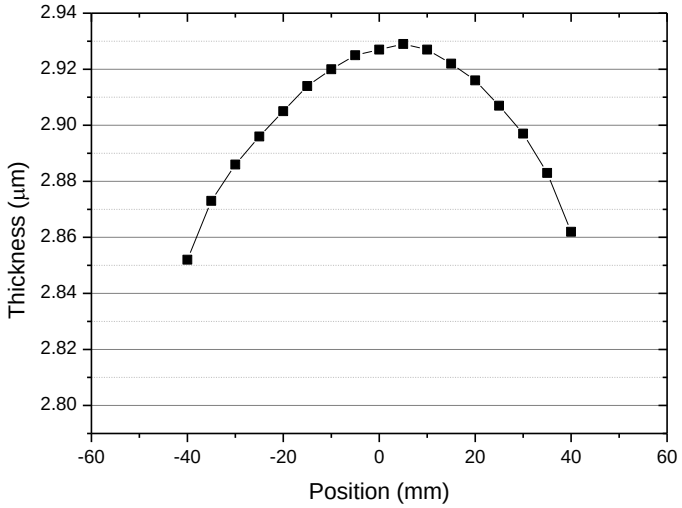


Figure 4.3: FTIR line scan across the wafer showing the thickness variation of an epitaxially grown Silicon layer.

To counter this variation in thickness across the wafers, samples chosen for analysis were cleaved from the centre of each wafer. As only small pieces are required for each experimental technique the limited number of central pieces was not an issue. This sample, shown in Figure 4.3, was chosen as the thickness was similar to that of a SPAD structure, and showed uniformity of 0.4% (+/-0.2% either side) over the 100mm diameter wafer. This was regarded as an acceptable level of variation and meant that samples chosen from anywhere within this central area would be representative of the wafer as a whole.

4.2.2 Growth of Germanium absorber region

Due to the 4.2% lattice mismatch the Germanium layer had to be grown via a two stage method [21]. The alternative approach would be to use long annealing stages post growth to reduce the TDD in the layer. However this would be unsuitable for the final device due to the doped layers surrounding the Germanium absorber. The annealing steps would cause large amounts of diffusion from these layers which as previously mentioned would disrupt the electric field profile.

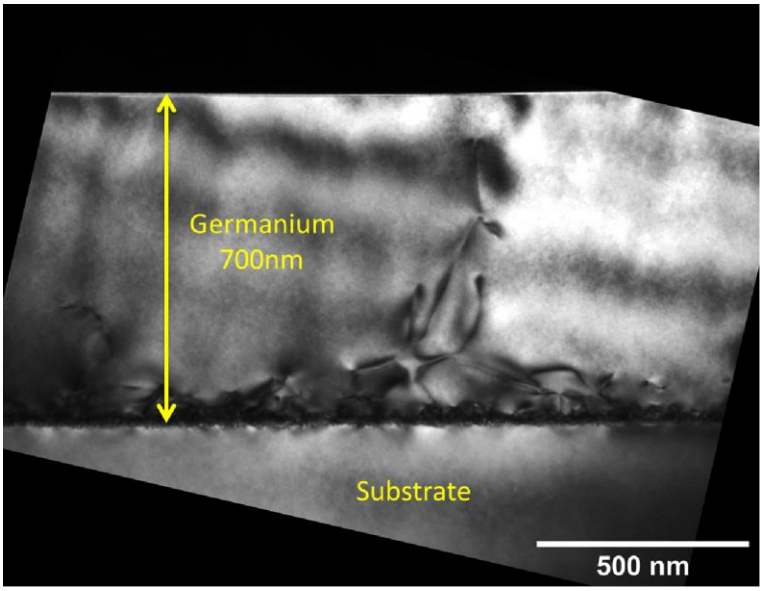


Figure 4.4: Dark field (220) TEM image of epitaxially grown Germanium with low and high temperature regions.The LT region is the highly defective layer at the interface level, whereas the HT layer is the subsequent region.

With the two stage growth process the Germanium layer was grown with a thin low temperature and a thick high temperature layer. The thin low temperature layer was grown at 600^{o}C and contained a very large TDD. This highly defective layer can be seen in Figure 4.4 around the interface between Silicon and Germanium. This layer was followed by the thick high temperature layer grown at 800^{o}C. Because the low temperature layer was grown past the critical thickness the layer was able to relax to that of bulk Germanium. Therefore when the high temperature layer was grown there was no lattice mismatch and no source for the formation of dislocations. A final annealing step was used to reduce the dislocation density. This was performed at 900^{o}C for 5 minutes. At this temperature and time interdiffusion should be kept minimal, while allowing for annihilation of dislocations. This allowed for the majority of the absorber layer to be relatively dislocation free as seen in Figure 4.4. HF etching was used to study the number of dislocations in the layer. The iodine etchant (containing HF) selectively etched areas where threading dislocations had terminated at the surface. A series of 12 images were used to give a fair representation of the structure as a whole. The TDD measured from the sample was approximately $5x10^{7}cm^{-2}$.

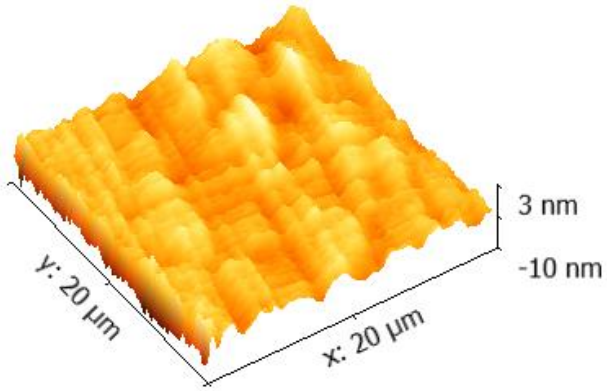


Figure 4.5: 3D AFM scan from the surface of Germanium on Silicon with HT and LT growth process. RMS 0.6nm. A cross hatch pattern may be observed on the surface corresponding to the misfit dislocation network at the substrate.

Figure 4.5 shows an AFM scan from this sample to ensure the surface was smooth to allow for precise fabrication. The RMS roughness of this sample was measured to be sub nanometre (0.6nm) which compares well to some of the best results from literature [21]. Annealing steps and LT/HT growth techniques are generally employed to reduce the roughness to sub nanometres, and Shah et al. predicted a thickness of over 900nm for the HT layer to have optimal RMS roughness.

The two stage growth had to be individually calibrated for the low and high temperature. To calibrate the growth rates a series of multilayer structures were grown at different temperatures. Silicon-Germanium layers were used as markers to separate each Germanium layer. The Germanium layers in each sample were grown at the same temperature and with a changing growth time. An example TEM image is presented in Figure 4.6. This allowed for a growth rate to be calculated for each temperature. The crystal quality of this layer is essentially perfect with no signs of any growth issues, which is vital as it is used for a platform on which to grow the high temperature layer.

An important requirement for a SPAD device is that there is no Silicon-Germanium alloy present in the structure. This is likely to reduce the overall effectiveness of the device as shorter wavelengths may be absorbed. While this is not problematic to the general device (operating at any wavelength) it is undesirable for optimising a SPAD operating at either 1330 nm or 1550 nm. Also thicker regions of Silicon-Germanium will directly lead to a reduction in the thickness of the absorption or

multiplication region as a result, which can lead to a decrease in absorption of the incident photon. The main parameter which affects the quality of the epilayer interfaces is the temperature at which the succeeding layer is grown. The higher growth temperatures will result in large amounts of interdiffusion (caused by the migration of Silicon and Germanium atoms) of the layers which can form an intermediate alloy layer.

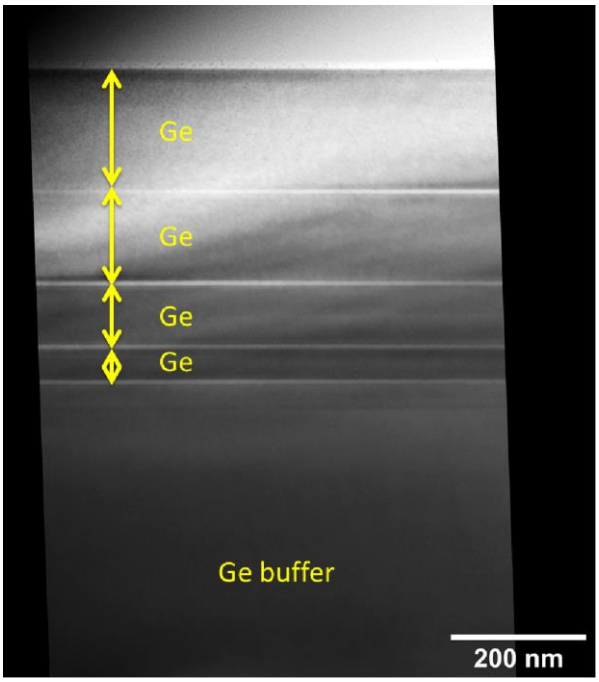


Figure 4.6: Straight through condition TEM image of Germanium and Silicon-Germanium multilayers grown at 350^{o}C. Thin strips are the Silicon-Germanium spacer layers.

From Figure 4.7 it can be noted that there is only significant intermixing of the layers when growth temperatures exceed around 700^{o}C (which is a typical high growth temperature used for SPAD structures). At these temperatures the diffusivity of the Silicon atoms in Germanium may be several orders of magnitude higher than other dopant atoms. Especially when long growth times are required, which is often unavoidable for SPAD structures, the potential for an alloy layer is increased. The diffusivity for Germanium in Silicon is substantially smaller, many orders of magnitude less. Therefore there is unlikely to be any noticeable diffusion from the Germanium absorber into the Silicon multiplier.

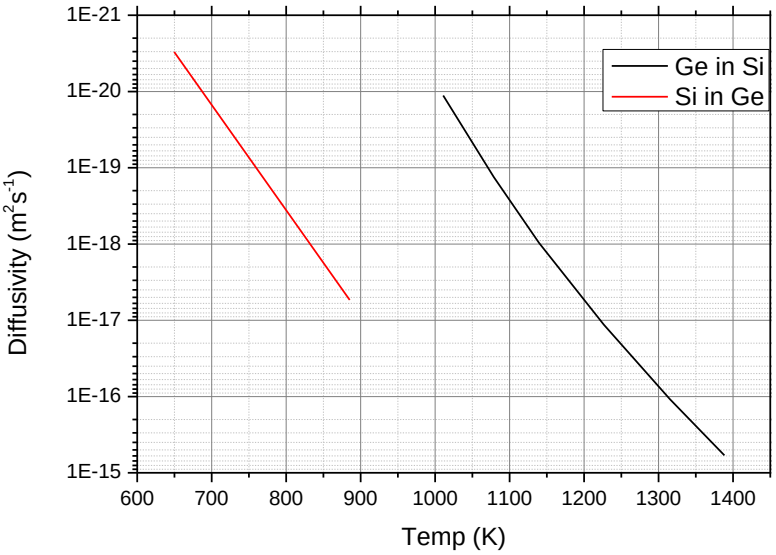


Figure 4.7: Diffusion coefficient for Silicon and Germanium interdiffusion against temperature. Adapted from [27].

Due to the nature of the two step growth (LT/HT) technique (section 4.4.2) used to grow the absorber layer, the growth temperature is conveniently low immediately after the charge layer, which is described in section 2.6.3, has been grown. As can be inferred from Figure 4.7 the diffusivity between the two layers should be negligible. However, growth of the higher temperature layer could cause some diffusion to occur.

The SIMS profile in Figure 4.8 highlights the minimal diffusion between the two layers. As can be observed the absorber region is essentially purely Germanium across the whole thickness of the layer, with no alloyed layer. Small amounts of interdiffusion can lead to quite significant increases in band gap, and therefore regions in the device which are unusable for applications at 1330 nm and 1550 nm. The level of Germanium diffusion into the Silicon layer is negligible, which is consistent with Figure 4.7. Therefore the only concern should be with Silicon or dopant atoms diffusing into the Germanium. There is a small amount of diffusion of Silicon into the Germanium over approximately 100 nm. It should be noted that diffusion of Silicon in Germanium is actually greater than many of the common n- and p-type dopants. While the sharpness of doping profiles may have a more

significant impact on the device (and is explored in the next section), optimised growth temperatures for the bulk materials plays an important role.

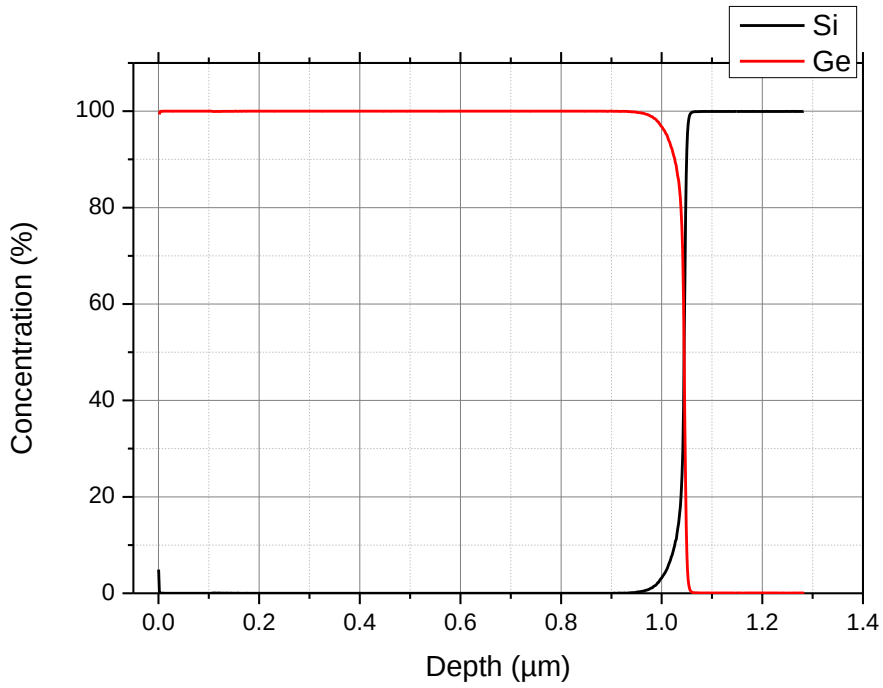


Figure 4.8: SIMS profile of a Germanium layer grown onto a Silicon epilayer with two step growth .Germanium layer is shown in red, while the Silicon is represented by the black profile. There is some intermixing at the interface.

4.3 Optimization of doped layers

The doping of a SPAD structure is of vital importance when ensuring optimal device performance. The precise doping of the layers dictates the electric field profile in a SPAD. Large diffusion tails or segregation of the dopant atoms will alter the electric field. During epitaxial growth, as well as other techniques including implantation, migration of these atoms is unavoidable completely. It has been essential to minimise any dopant segregation and diffusion through the choice of dopant and growth technique employed. The most popular p-type dopant is boron and the most popular n-type dopants are phosphorus and arsenic. Samples were grown using each of these dopants and have been measured, predominantly through SIMS profiling. The standard substrates used for growth were doped with Arsenic to a concentration of $5x10^{19}cm^{-3}$. This was ideal for the SPAD structures, so initial designs were grown using the substrate as the highly n-doped layer.

4.3.1 Segregation in the layers

Many of the samples analysed during the optimization period of this work were done so simultaneously with the growth of SPAD structures. Initial work looking at the substrate and Silicon multiplication region interface was performed on a SPAD to observe potential problems which may occur in a final device.

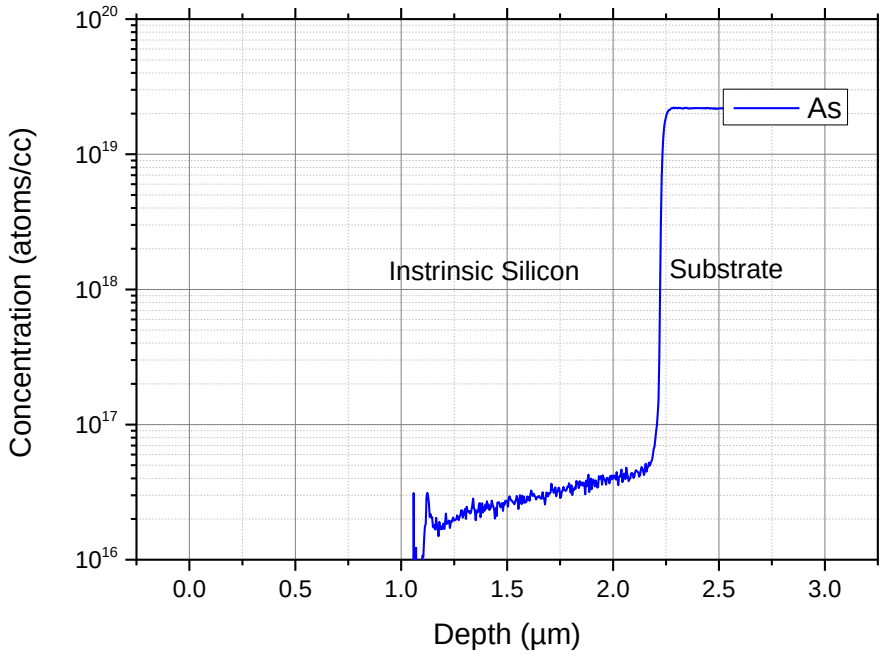


Figure 4.9: SIMS profile of arsenic concentration from substrate through into Silicon epilayer. Arsenic concentration segregating into the epilayer can be seen with the decreasing tail in the SIMS profile.

SIMS analysis of arsenic concentration is shown if Figure 4.9, where the substrate is at a depth of approximately 2.25 μm, which is seen as the abrupt straight line. A small level of diffusion can be seen from the substrate as the line is not perfectly vertical. Despite this, the diffusion from the substrate is only 11nm/dec.

However n-type dopants are well known for having a high surface affinity [86], and therefore there is a fairly significant level of segregation from the substrate. Also local strain fields are associated with larger (or smaller) atoms sat in lattice sites which can result in them being forced out and therefore upwards in the layer.

This is characterized by the shallow extension of the doping profile into the Silicon epilayer. Unlike the sharpness of the diffusion tail the segregation was measured to be 2.59 μm/dec. This provides a potential issue for the SPAD multiplication region. The layer should be nominally undoped, but it is clear that due to the segregation, instead on having a doping concentration below $10^{15}cm^{-3}$ described in the growth plan, the concentration of n-type dopants in the layer is over an order of magnitude higher. The nature of the dopant does mean that generally any layer succeeding an n-doped region will experience some extra doping through the segregation process. The potential improvements in preventing some of this segregation are discussed in more detail later in the chapter.

4.3.2 Reducing segregation in the multiplication layer

The segregation through a layer sees atoms move up towards the surface of the structure, often travelling beyond their desired position. This becomes problematic when regions of intrinsic material become lightly doped by a neighbouring layer. As is shown during this section, the concentration of foreign atoms taking residence in other layers may be several orders of magnitude lower than their corresponding layer. Despite this, the concentration observed is still in excess of the generally accepted intrinsic level of doping ($\approx 10^{15}cm^{-3}$). As segregation is associated with the surface, it is unlike diffusion in the sense that the tails observed in the profiles are much longer. Diffusion tails can be seen to fall away to the background limit much faster than those linked with segregation, whose tails can be seen extending several microns through the succeeding layers.

Another, potentially more serious issue, especially when considering the future devices from this work, is the smearing of the dopant profile, similar to that seen for diffusion. Reducing the amount of segregation, particularly for n-type dopants, requires different growth techniques to a standard one temperature method. As was demonstrated in Figure 4.9, even growth from a substrate will experience a large amount of segregation, at this high temperature.

It should be noted that segregation is inevitable to some degree when growing epitaxially. Before growth of a layer which is preceded by a doped layer may be initiated the gases in the system must be removed to ensure that no dopant gases are present when growing undoped material. Unavoidably the wafer will still be

subject to the growth temperature in the chamber, leading to dopant atoms rising to the surface. While some diffuses downwards the contribution upwards from diffusion and segregation is more significant.

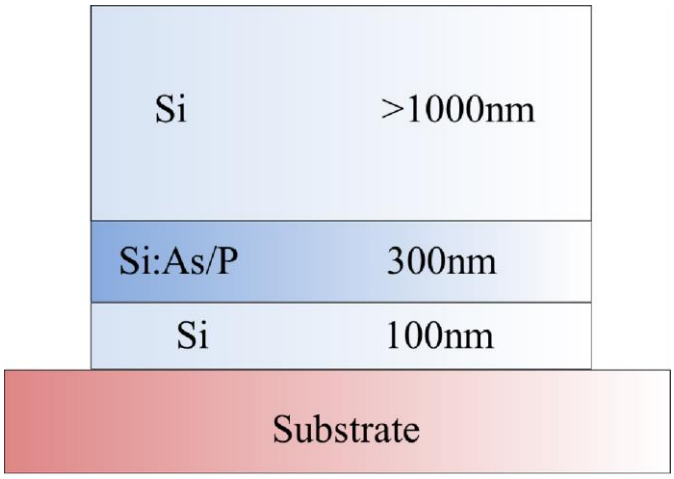


Figure 4.10: Growth plan for arsenic doped Silicon layer grown on 100nm of intrinsic silicon, with 1µm of succeeding intrinsic silicon. The 300nm doped layer can either be doped with As or P. The central doped layer was grown at 800^{o}C, and the top Silicon layer was grown at 900^{o}C.

An arsenic doped Silicon layer was grown according to the growth plan described in Figure 4.10. The doped layer was grown at a temperature of 800^{o}C, whereas the Silicon layer grown on top used a temperature of 900^{o}C. This high temperature growth of the bulk layer has contributed to a large segregation tail which can be observed in Figure 4.11. This sample demonstrates the amount of segregation while growing under standard conditions. The segregation tails were measured to be 2.45μm/dec, and can be seen to have arsenic concentrations as high as 10^{18}cm^{-3} after 1 μm of undoped Silicon growth. This doping concentration is comparable, even exceeding that of, the charge sheet in the SPAD structure. Clearly this high level of As concentration is unsuitable from a device perspective. The multiplication region would then not be intrinsic and there would be no constant high electric field through the layer, which is essential for a successful device.

The alternate n-type dopant commonly used in Group IV epitaxy is phosphorus. A similar structure was designed for analysis, this time consisting of a phosphorus doped silicon layer. Again the doped layer was grown at a temperature of 800^{o}C, whereas the Silicon layer grown on top used a temperature of 900^{o}C. Growth of the phosphorus

and arsenic doped samples was under the same conditions to compare the level of segregation. Figure 4.11 shows the SIMS profile of the phosphorus sample. It is immediately clear that there is significantly less segregation from the layer. Even using a high growth temperature for the intrinsic layer yields a relatively modest segregation tail of 0.5 μm/dec. Over roughly a micron the doping concentration for phosphorus has dropped by over an order of magnitude, whereas for the sample doped with arsenic over the same range the doping concentration has only reduced by a factor of two. Due to the smaller atomic size of phosphorus fewer atoms have been forced upwards during growth.

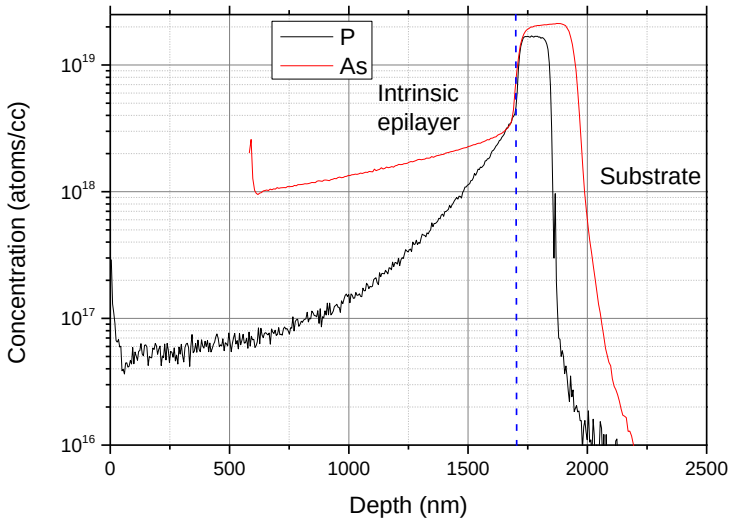


Figure 4.11: SIMS profile comparison of arsenic doped and phosphorus doped Silicon. The detection limit is around $10^{14}cm^{-3}$ for this measurement.

To investigate the reduction in segregation a similar structure was grown, Sample 14-111. The growth plan is described in Table 2, where the only difference is seen immediately after the doped layer. The initial 30 nm of the intrinsic Silicon was grown at a reduced temperature of 600^{o}C. Although impractical for a full layer, a lower temperature, and therefore slower growth time, layer could reduce the level of segregation. To grow the multiplication region all at this temperature would not be viable when mass producing the SPAD structures, so subsequent growth resumed at the higher temperature (900^{o}C).

Table 2: Growth details for each structure grown. The sample numbers along the top are used to identify each sample. 14-112 is the original layer without any additional techniques used.

	14-111	14-112	14-113	14-114
Top Layer	1000nm 30nm LT	1000nm	1000nm Etch then 30nm LT	1000nm Etch
Doped layer	200nm	200nm	200nm	200nm
Bottom Layer	100nm	100nm	100nm	100nm

At this stage several other methods to reduce the segregation were explored. The first of which was the addition of an etch step. This was performed straight after the doped layer growth, also described in Table 2. The final intrinsic layer growth was performed at 900^{o}C so that it could be compared to the original high temperature layer from Figure 4.11. This was sample 14-114.

The etch step was used to remove a small layer from the phosphorus doped Silicon. The purpose of this step was to remove a small layer of the surface where there could be a build-up of phosphorus atoms undergoing segregation. A result of this would be a reduced layer thickness, therefore if implemented into a SPAD structure; the doped layer would be required to be slightly thicker. The issue with having to account for this loss in thickness is that there would have to be extra calibration to determine exactly how long the etch step should be to result in a final layer with the desired thickness. Therefore, this technique would need to produce a distinct improvement to warrant further consideration.

A final sample, 14-113, was grown which combined the low temperature intermediate layer with the etch step, shown in Table 2. The growth followed the same routine as before for each stage, which consisted of the etching, followed by 600^{o}C and 900^{o}C steps.

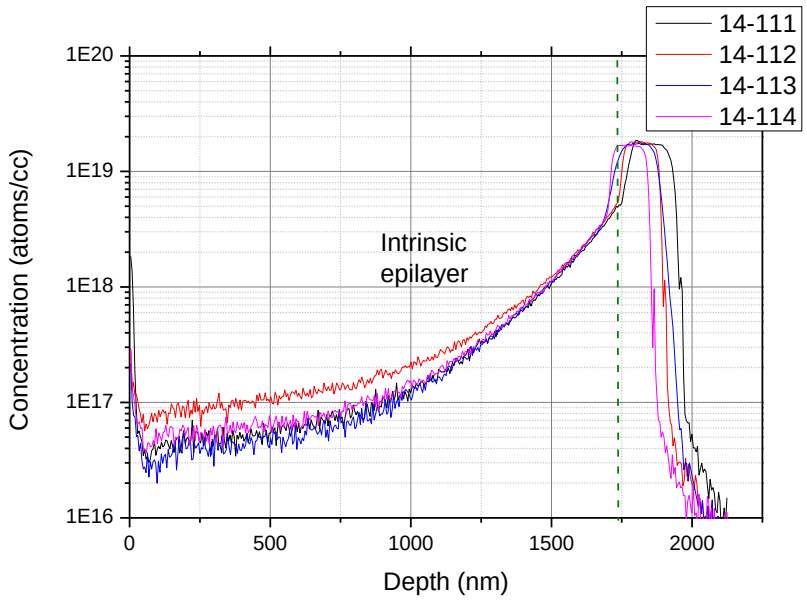


Figure 4.12: SIMS profiles of each phosphorus doped silicon sample. Diffusion tails can be observed protruding into each intrinsic silicon region. The doped layers are the rectangular regions in each profile.

The SIMS data of sample 14-111 in Figure 4.12 shows an improvement in the segregation from the doped layer. The very small bump immediately before the intrinsic layer is thought to be caused by some surface accumulation during the cooling of the temperature. The tail associated with this profile was measured to be 380nm/dec, therefore confirming the benefits of using this growth technique. As the temperature is lower, there will be less energy for the atoms to be able to move, which will lead to atoms becoming "locked in" to lattice points more readily. If the layer is able to be fully covered the dopant atoms will not be able to move up through the layer.

As can be observed in Figure 4.12 the small bump prior to continued Silicon growth has disappeared with the etch process, in sample 14-114, as anticipated. The segregation tail for Sample 14-114 was measured at 360 nm/dec, which is a slight improvement of the method used for Sample 14-111. This is thought to be caused by the removal of any surface accumulation of phosphorus atoms.

The SIMS profile for sample 14-113 has a measured segregation tail of 425nm/dec. Similarly to sample 14-114, the small bump at the interface between the doped layer and the intrinsic layer has been removed due to the etching process.

Interestingly the gradient of this tail is larger than both other techniques involving the low temperature buffer or the etch. This could potentially be caused by more P atoms becoming exposed during the etching process. If this is the case when the LT growth is resumed there could be a clustering of different atoms, as there is less energy on the surface. This could then lead to P atoms more readily segregating as the surface is not completely covered by Silicon atoms.

The results for this investigation of the segregation tails are tabulated in Figure 4.13. The most significant improvements were made when using techniques with just an etch or just the low temperature buffer layer. The segregation observed in sample 14-114 shows an improvement of nearly 7 times over the arsenic layer.

	Comments	Segregation
14-111	LT Si layer after Si:P	378nm/dec
14-112	HT Si layer after Si:P	506nm/dec
14-113	Etch and LT Si layer after Si:P	425nm/dec
14-114	Etch then HT layer after Si:P	360nm/dec

Figure 4.13: Table of segregation tail results.

A drawback to using the etch approach is that it can be difficult to know how much material is removed. Extensive calibration would be required to calculate the thickness which would be etched. This can be problematic for strict growth plan thicknesses. It was therefore decided that further improvements would be attempted using the low temperature layer method (as in sample 14-111).

4.3.3 Reduced temperature and multiple LT/HT steps

To investigate the low temperature buffer layer technique several options were considered. The first was to grow a reduced temperature low temperature layer, which was essentially grown according to Table 2. This was sample 14-314, where the only differences in growth plan was the temperature of this growth (500^{o}C).

The segregation tail from the SIMS profile of 14-314, shown in Figure 4.14, was measured to be 420 nm/dec. Most noticeably at this reduced temperature, the concentration of phosphorous in the nominally intrinsic layer immediately after the

doped layer is approximately an order of magnitude lower ($10^{18}cm^{-3}$ compared to $10^{17}cm^{-3}$ in sample 14-314).

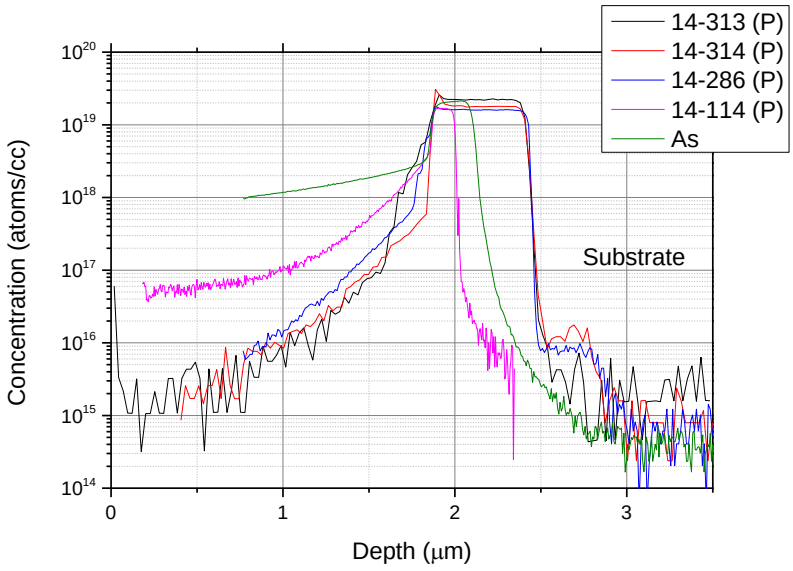


Figure 4.14: SIMS profiles for each phosphorus sample using lower temperature growth, along with LT/HT repeats. A comparison to the As doped layer is also shown.

However there is an elevated concentration at the surface of the doped layer, in the form of a bump, which is similar to sample 14-111. For 14-314 the bump appears to be more pronounced. This is probably caused by the extra time required to reach the lower temperature of 500°C, which has resulted in more phosphorous accumulating at the growth surface.

As the low temperature layer was grown at a lower temperature less phosphorous was able to segregate through the epilayer during growth, which is attributed by a maximum concentration of approximately $5x10^{17}cm^{-3}$. Also the surface concentration of phosphorus does not contribute as much phosphorus to the epilayer due to the low temperature growth. Consequently after half a micron the concentration in the intrinsic region is three orders of magnitude lower than the doped layer. Although this is still above the ideal $10^{15}cm^{-3}$ there is a significant improvement upon even sample 14-111. A drawback to simply reducing the temperature much further is the increased growth time which would be necessary, and the activation of dopant atoms.

Building on this improvement, an investigation into multiple repetitions of the low temperature layer was performed. Initially a sample was designed with just two repetitions of low temperature layers. After the initial Silicon-phosphorous layer was grown, the normal low temperature layer was deposited at 500^{o}C. This was followed by a thin high temperature layer grown at 800^{o}C, before a second low temperature layer was grown at 500^{o}C. Finally the structure was completed with a thick layer of Silicon grown at the higher temperature.

The segregation tail for sample 14-286, described in Figure 4.14, has a gradient of 420nm/dec, which is the same as sample 14-314. As this technique utilises the same method as the single low temperature sample, this is unsurprising. The initial concentration in the intrinsic layer is slightly higher than sample 14-314, measured at approximately 7x10^{17}cm^{-3}. This potentially could be caused by an increase in segregated atoms in the low temperature layer during the thin high temperature layer and subsequent cooling time. It should be noticed that there is a slight increase in thickness caused by the step growth processes.

A third sample was grown with multiple repeats of the alternating low and high temperature layers. The purpose of this sample was to investigate any potential reduction in the initial dopant concentration in the Silicon layer. Design of this sample (14-313) followed the same recipe as sample 14-286, however there were six repeats of the thin low and high temperature layers. As expected the segregation tail is approximately 420nm/dec. The error on this measurement is +/-80nm/dec due to the high noise level approached soon after the interface between the final low temperature layer and the intrinsic layer.

There are two key findings from this final sample, which provide both benefits and drawbacks to this multilayer technique to reduce the n-type segregation. Probably the more noticeable point is the added thickness to the layer, albeit with decreasing concentration. Obviously there is an extra thickness to the layer, which is currently around 300 nm, but could potentially be mitigated by reducing the thickness of each high and low temperature step.

Despite this possible problem, the initial phosphorous concentration in the intrinsic Silicon layer is less than 10^{17}cm^{-3} which demonstrates an advantage over samples 14-313 and 14-286. Also the phosphorous concentration is of the order of 10^{15}cm^{-3}

after only half a micron, which is equates to a drop of four orders of magnitude over a total thickness of less than one micron.

It can be seen that there are advantages and disadvantages of each method, which would depend on the application. Without using any multiple stage process the result is a more abrupt doping profile, with a slightly higher initial dopant concentration. Using two repetitions of low and high temperature steps, the result is a slightly less abrupt doping profile and an initial concentration comparable to that of the single high and low temperature layers. Multiple repetitions have a much less abrupt doping profile, but have the lowest initial doping concentration in the intrinsic layer. When comparing the two extremes (14-313 and 14-314) the concentration at the same depth is actually quite similar, but there is an improvement for sample 14-313 by about a factor of two over 14-314 (visible in Figure 4.14 at 1.5um depth).

Table 3 quantitatively describes, and summarises the data for each sample regarding the doping concentration in the intrinsic material. This is designed to concisely demonstrate two of the main advantages/disadvantage of each technique. As previously mentioned the etching technique used in samples 14-113, and 14-114 would need extra calibration to account for thickness loss, but do show potential which could be further explored. As for the other samples it is unanimously clear that a reduction in growth temperature is beneficial, however when these steps are repeated the suitability is dependent on the device requirements.

There appears to be three competing processes throughout these samples: namely dopant migration, the locking in of atoms, and the evaporation of hydrogen. The third of these only really has any effect at the lower temperature samples as hydrogen evaporates from the surface above 500^{o}C (therefore acting as a surfactant). At the lower temperatures atoms have less energy to move through the lattice but clustering may become energetically favourable. However, at higher temperatures the dopant atoms are more likely to be locked into their positions due to the faster growth rate, but also will be subject to more dopant migration. Initial low temperature growth (500^{o}C) followed by LT/HT steps appear to give the most promising results. It should also be noted that any contribution from autodoping

(from the chamber walls) will also be reduced with either a low temperature or short growth time.

Table 3: Table describing the dopant reduction in the intrinsic silicon layer for each sample in the investigation. Peak concentrations were taken from the SIMS profile immediately adjacent to the doped layer.

Sample	Peak conc in intrinsic at/cc	Percentage reduction %	Extra thickness nm
14-111	5.30E+18	69.5	0
14-112	6.20E+18	65.6	0
14-113	3.70E+18	78.7	0
14-114	4.90E+18	71.2	0
14-314	5.00E+17	97.2	0
14-286	7.00E+17	95.9	120
14-313	9.10E+16	99.6	300

4.3.4 Doping concentrations for SAM SPAD device

As the SPAD is based on a pn junction, precisely doped layers are of the upmost importance. For the device to function effectively doping concentrations must be accurate and be subject to minimal diffusion. If this is achieved the electric field profile throughout the structure, whichever design is utilised, should be of the optimal strength in each region. For the separate absorption and multiplication region design this is essential, as the functionality of the SPAD is dependent on a low field in the absorber region and high field in the multiplication region. This is realised through a *pipin* structure where the charge sheet (the middle *p*-layer) provides a shield against the field in the *in*-region. For device operation the field will punch-through the charge sheet, but it is important that this does not occur before breakdown. In this case carriers generated in the absorber region will not be swept into the multiplication region. It is therefore apparent how smearing or incorrect doping profiles can result in a poor device.

Unfortunately the doping of semiconductor materials through the CVD technique is not straightforward, especially when atoms are prone to deviation from their original location through diffusion processes. Calibration is an extremely important aspect of optimising SPADs and must be performed for each dopant with its respective host lattice. The main doping combinations used in the SPAD designs which are epitaxially grown (i.e. not a doped substrate) are: Ge:B, SiB, and Si:P.

A series of multilayer samples were grown for each of the dopants with varying ADC values. ADC values are a variable which controls the flow of dopant gas relative to carrier gas, and are controlled through the CVD. The multilayers were designed as alternating layers of undoped and doped material. Each doped layer had a different ADC value so that the doping concentration of each doped layer increased towards the surface of the structure. The range of doping concentrations was selected to be of the order $10^{16}cm^{-3}$ to $10^{19}cm^{-3}$ as this was the range of doping values used for the SPAD devices.

The first sample was 14-292, a boron doped Germanium multi-layered structure. To analyse the doping concentration and the diffusion from each layer a SIMS profile was measured and is presented in Figure 4.15. Growth temperatures used mimicked that of the doped layers in the final SPAD structure (600^{o}C).

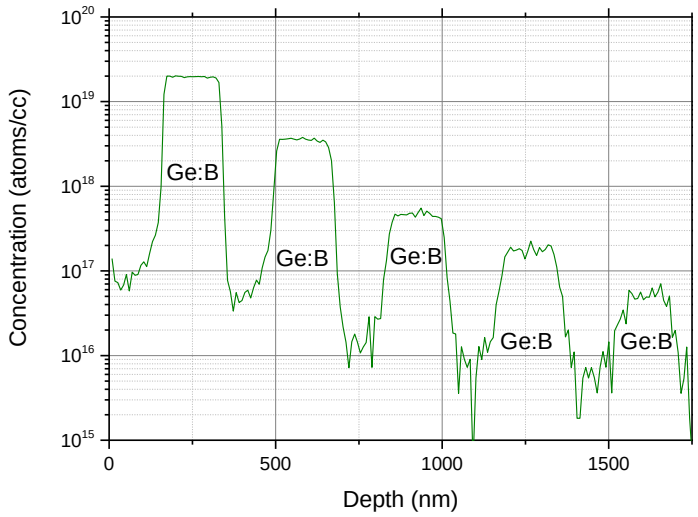


Figure 4.15: SIMS profile of Ge:B sample 14-292. Layers were grown as alternating doped and undoped layers, with each doped layer having an increased concentration.

Abrupt doping profiles were obtained for each layer, with the majority of diffusion tails at 15nm/dec or less. Deeper layers, which would have been subject to higher temperatures for a longer period of time throughout the rest of the growth, had slightly higher diffusion tail (around 30nm/dec). When growing these layers in a real device they would not be subject to as long a growth times.

By plotting the ADC values against the doping concentrations for each layer a direct relationship between the two can be established, and therefore ADC values can be selected accordingly.

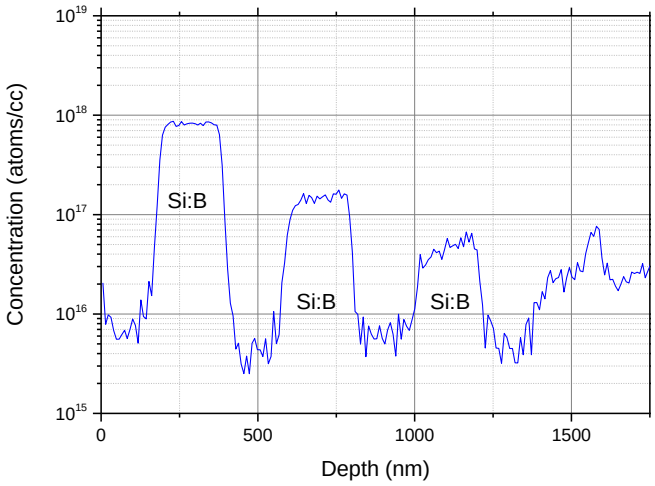


Figure 4.16: SIMS profile of Sample 14-291.The layer consists of alternating doped and undoped layers of Silicon and boron doped Silicon respectively.

Sample 14-291, was grown with the same approach as 14-292. This sample contained Boron as the dopant atom, and Silicon as the bulk material. SIMS analysis, shown is Figure 4.16, showed a series of doped layers with concentrations in the region of $10^{16}cm^{-3}$ to $10^{18}cm^{-3}$. This data was selected because the charge sheet layer in the separate absorption and multiplication SPAD required a lower doping concentration than the other p and n region. For other designs a similar structure was used to 14-292, with appropriate doping concentrations.

For samples which required higher levels of boron concentration a structure with higher doping was utilised, but is not shown here due to its very limited use in the growth of SPAD structures.

4.4 Single photon avalanche diode designs using Si and Ge

There is a wide range of choices for the design of a SPAD device. Most of the earlier SPAD designs relied on just one bulk material, or even one active region. Several different designs had to be considered when deciding on the structure for the SPAD device. A selection of the initial growth plans are presented in this section.

4.4.1 Pure Silicon SPAD design

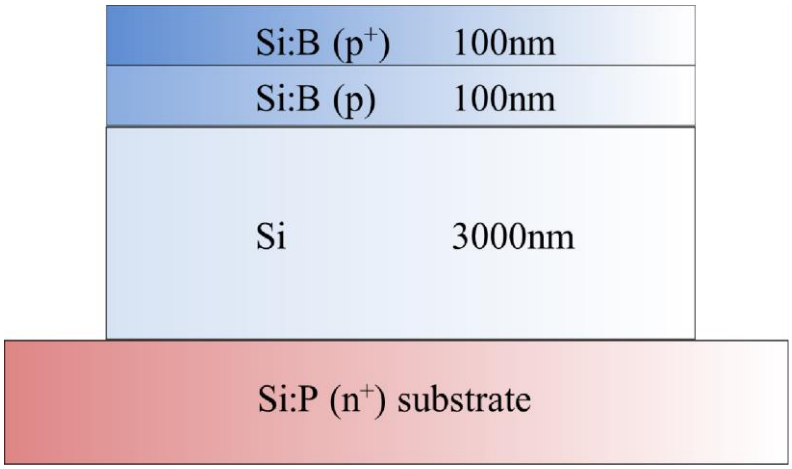


Figure 4.17: Growth plan for SPAD structure based purely on Silicon. A pin structure is used for absorption and multiplication of incoming photons. Sample ID is Si SPAD.

One of the original designs for the SPAD in this work was a simple *pin* structure. The thick intrinsic layer was used as an absorber and multiplier, shown in Figure 4.17. The thickness of the layer would give a relatively high probability that the incoming photon would be absorbed, but also incurred a long transit time for the photo generated carriers. At longer wavelengths even very thick layers would not be enough to effectively absorb a photon. Due to the presence of only one intrinsic region the high field will be across here, and there is no low field region for drift of carriers. This design uses Silicon throughout, but obviously for application to longer wavelengths Germanium would be more appropriate. Even with the thick layer for absorption photons with energies into the infrared range would have an exceptionally low probability of absorption.

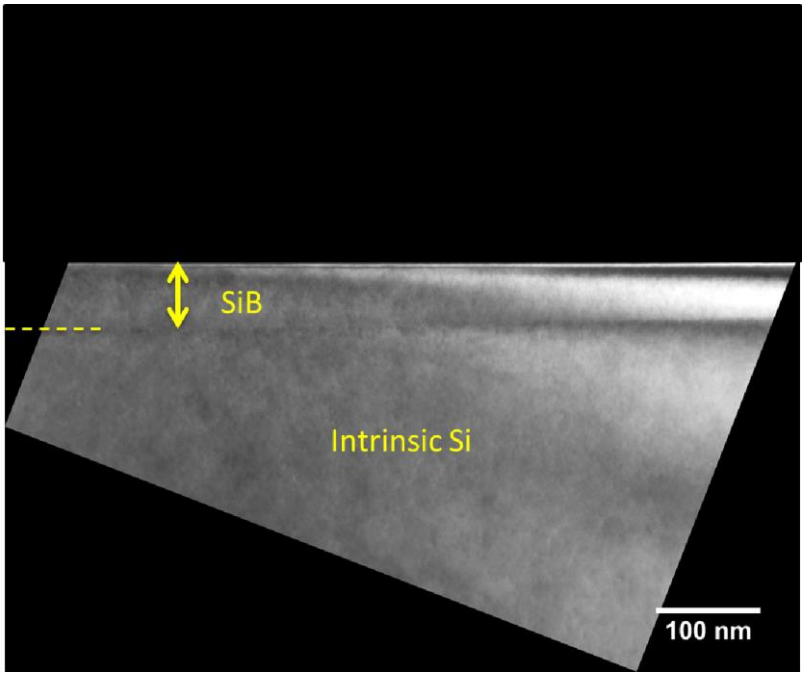


Figure 4.18: Dark Field 220 diffractive TEM image of top doped layer of Silicon SPAD design (Si SPAD). Highly doped top contact layer can be observed due to the growth stop (dark line).

Structural characterization of the surface and top contact layer is shown in Figures 4.18 and 4.19. The TEM image demonstrates the heavily p-doped top layer in the structure. The layer below which was lightly p-doped is not so clearly visible due to the minimal contrast between that and the intrinsic Silicon. The thickness of the top layer was measured to be 97+/-3 nm, which was consistent with the design of the structure.

An image of the full structure in Figure 4.19 shows no issues with the substrate/epilayer interface. Due to the lack of contrast, caused by a thicker region of material formed during the preparation (which was not electron transparent) it was necessary to estimate the interface location, which is shown by the red line in Figure 4.19. A large portion of the substrate was included to ensure that there were no issues in the case of a deviation from the expected thickness. Due to the difficulty in detecting the substrate doping level SIMS analysis was used to investigate the substrate.

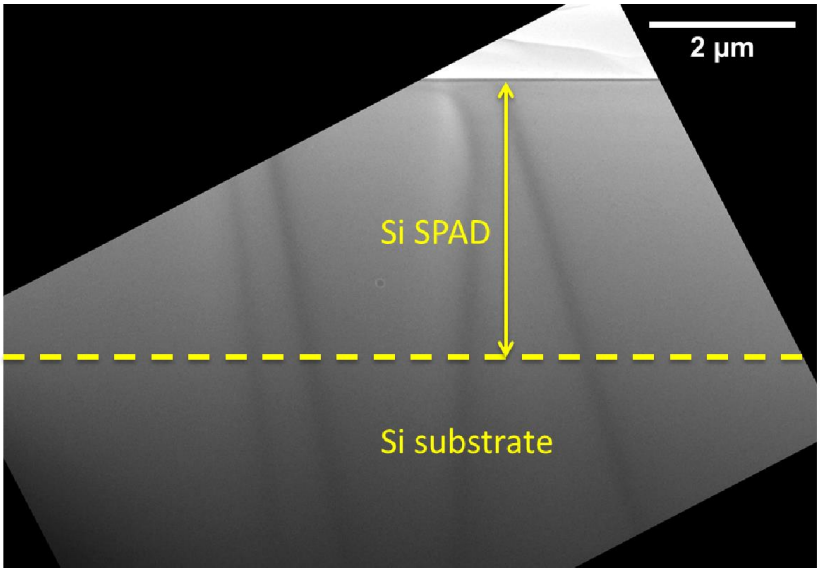


Figure 4.19: Straight through TEM image of full Silicon SPAD structure (Si SPAD).

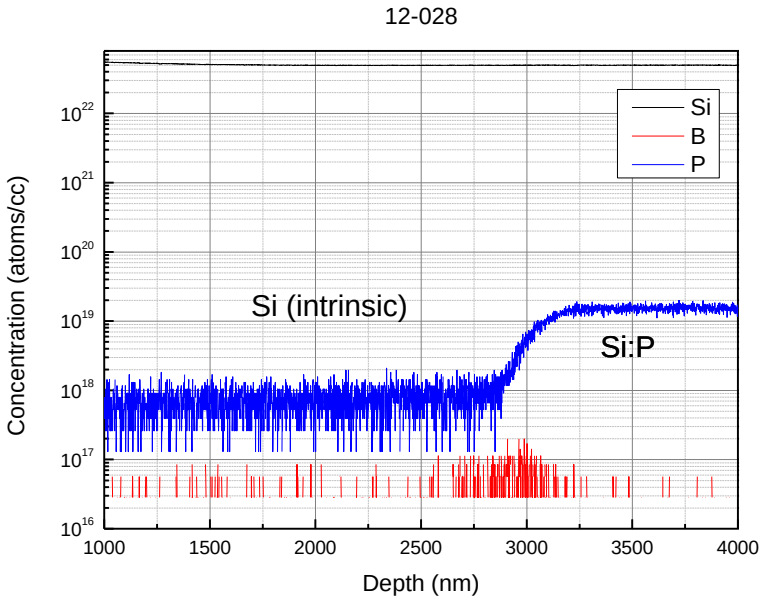


Figure 4.20: SIMS plot of the Silicon SPAD measured into the substrate (Si SPAD). Doping concentrations can be observed for the doped and intrinsic regions.

Figure 4.20 reveals that the phosphorus doping does begin at the 3μm mark, which indicates the doped substrate used for growth. Due to the high growth temperatures used to obtain a relatively thick layer there is some diffusion of the dopant from the substrate (250 nm/dec). However simply reducing the temperature for growth

would limit the potential for mass production because of the extensive growth time required. In any case the thickness of the intrinsic layer limits the effectiveness of this structure for device production.

4.4.2 Pure Germanium SPAD design

An alternative design is that of an all Germanium SPAD. The advantage of using this as a device would be for application to infrared wavelength operations. Similarly to the all Silicon SPAD this structure is essentially *pin*, therefore absorption and multiplication occurs in the same region. The growth plan is presented in Figure 4.21.

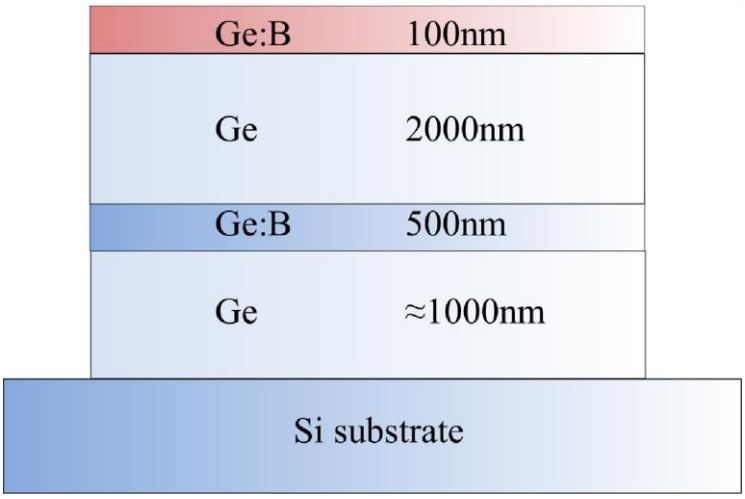


Figure 4.21: Growth plan for SPAD based purely on Germanium. The bottom intrinsic layer grown initially is used as a buffer to minimise TDD in the active region of the device (pin area). Sample ID is Ge SPAD.

A noticeable difference to that of the all Silicon SPAD design is the additional intrinsic layer succeeding the lightly doped substrate. This plays no part in the operation of the SPAD, but acts as a buffer layer. This region will be full of dislocations due to its mismatch with the Silicon substrate, but will relieve all the strain and should ensure that the devices layers are of good quality. This approach allows each layer to be grown at higher temperatures, without the need of a low and high temperature technique. A draw back to this is similar to that of the InGaAs/InP structure. When incorporating these onto cheap Silicon substrates this extra buffer presents a similar disadvantage to that of a Silicon Germanium alloy buffer used to overcome the lattice mismatch between the III-V material and

Silicon. Despite the improvement in the TDD in the active region, there will still be a number of dislocations as well as a potentially rough interface, which could affect the further growth quality.

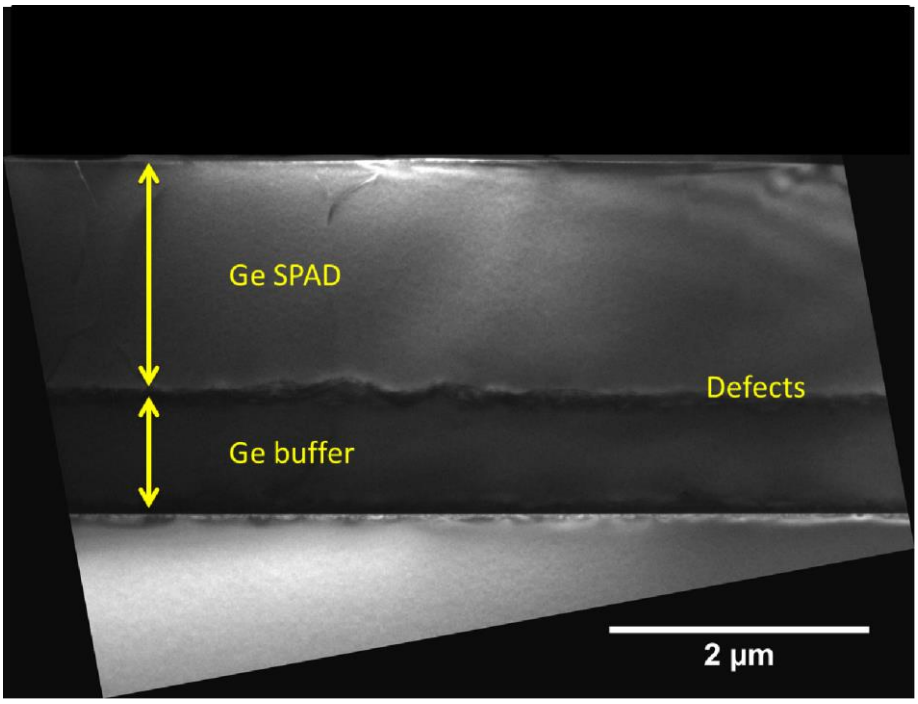


Figure 4.22: Cross-sectional TEM image of all Germanium SPAD structure taken under dark field (004) conditions. Ge SPAD.

The TEM image in Figure 4.22 clearly shows the interface between the buffer layer of Germanium and the *pin* region. The interface is very rough, which is most likely caused by the higher growth temperature and the subsequent defective layer. The roughness from the interface is present also on the surface on the structure seen in an AFM. TEM imaging has been beneficial in identifying the likely cause and the location of this growth issue.

The doped layers were identified using XRD, as they were not visible in the TEM image. A shoulder to the Germanium peak can be observed in Figure 4.23 representing the doped layer. The Silicon peak is observed on the right of the rocking curve, and the Germanium on the left. The general broadening of the Germanium peak is caused by the high TDD present particularly in the buffer layer, while the asymmetry (i.e. the shoulder) a sign of the highly doped layers. Due to the existence of both boron and phosphorus in the p- and n-layers a

concentration cannot be calculated, however this does show their presence which was not seen in TEM, through lack of contrast.

Potential introduction of low and high temperature regions to the Germanium buffer layer could improve this structure. This technique is used in the future structures discussed in the next section. It was preferred to introduce this method into the separate absorption and multiplication SPADs due to the issues about incorporation onto Silicon substrates discussed above.

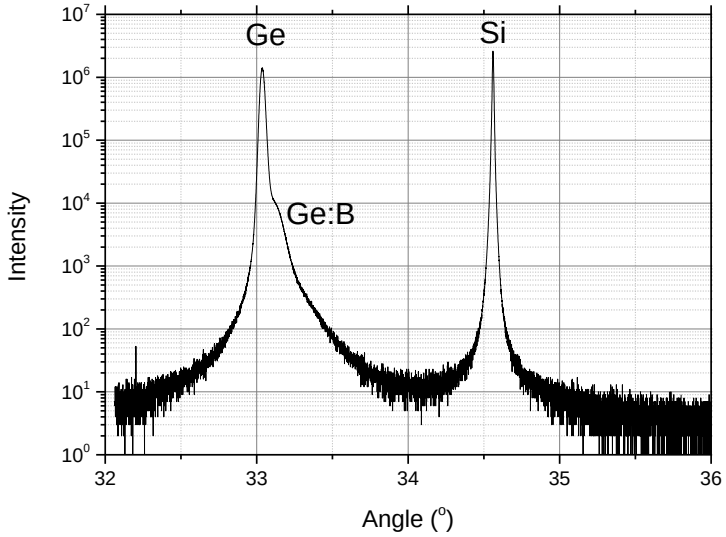


Figure 4.23: Rocking Curve of the all Germanium SPAD. The doped germanium layer can be seen on the side of the germanium peak. Ge SPAD.

The most common SAM SPAD design is considered probably the most effective design for a SPAD, with the majority of designs in literature utilising this structure. The thickness of the various layers may be altered, ranging from one to several micrometres. Thicker absorption regions are more likely to absorb the incoming photons, but the device response is likely to be slowed by the extra time the photocarrier spends traversing to the multiplication region. Multiplication regions should be thick enough to ensure multiplication, but also shouldn't have long transit times. Therefore a compromise is required.

4.4.3 Summary of Si SPAD and Ge SPAD

In summary the all Silicon SPAD design holds some promise for application using shorter wavelengths than the infrared ones studied in this work. A main drawback when using a purely Silicon SPAD at longer wavelengths is the thickness of the layer to ensure absorption of the photon. This will cause the device to be slow. However the lack of dislocations throughout the layer is beneficial in reducing the dark count rate for the device. Highlighted here is the superb quality of the p-doped layer used as the top contact, as well as the overall quality of the layer, even if the devices capabilities are limited to higher energy photons.

Silicon has a long absorption length, and even at visible wavelengths around 500 nm the layer would need to be around two microns thick. For infrared wavelengths (e.g. 1.3 or 1.55 um) the layer thickness would have to be impractically thick. Obviously this is impractical for a device where the transit time of photogenerated carrier is important. This is why the previous all-Silicon design had a thicker layer than when using Germanium. Ideally this device would still have a respectable probability of photon absorption around 1μm, as well as a suitably low time for the carriers to traverse the layer. Germanium has a much superior absorption length when compared to Silicon, especially at longer wavelengths. For infrared photons at 1.3μm, which are desirable for this work, the absorption length is as little as 1μm. Importantly this is comparable to the III-V detectors which operate at the same wavelengths [87]. The SPAD designed using only Germanium holds more promise for these wavelengths, but further optimization would be necessary to try and avoid problems associated with the highly defective regions. However, as Germanium suffers from higher excess noise, future efforts were concentrated towards a separate absorption and multiplication region SPAD structure.

4.4.4 Ge and Si SAM SPAD designs

Figure 4.24 describes one of the early separate absorption and multiplication region SPADs which has become the basis of the final structures designed and characterized in this work. As discussed in Chapter 2 this design operates by absorption of a photon in the narrow band gap Germanium, subsequent traversing

by the electron into the high field Silicon layer, and impact ionization of carriers leading to a runaway avalanche current.

This sample was grown before much of the optimization work, and therefore demonstrates several potential performance damaging problems, which has been included to highlight the structural issues of a non-optimized device. The complete SIMS profile is presented in Figure 4.25, which is seen to follow the general growth plan from Figure 4.24.

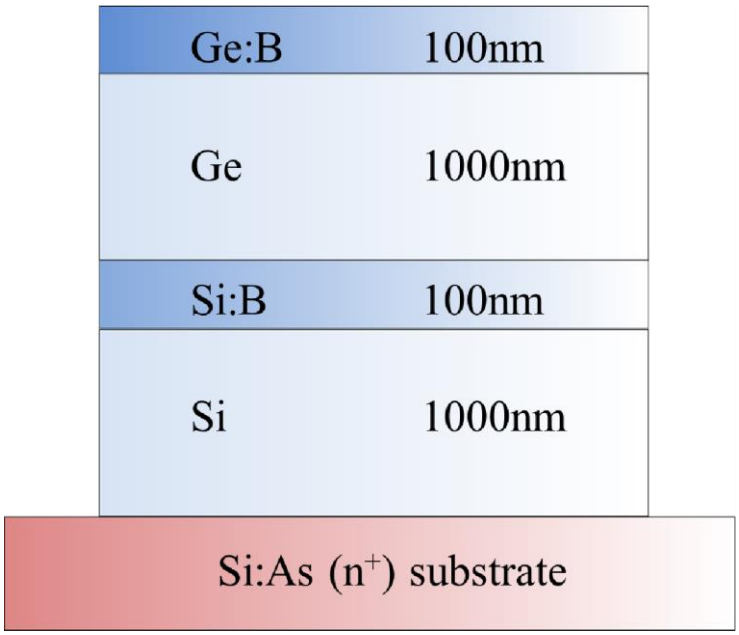


Figure 4.24: Growth plan of separate absorption and multiplication region based on Silicon and Germanium. Sample ID SiGe SPAD1.

However, despite following the desired growth plan, there are several factors which present concerns for SPAD or even photodiode operation. The dopant levels for both the charge sheet and the contact layer are not uniform. This is probably due to diffusion of the dopants at higher temperatures and a non-optimized ADC calculation. There may also be some clustering of dopant atoms which has led to this spike. It is likely to disrupt the electric field profile which could result in no punch-through of the device, as the doping concentration appears much higher than planned. The spike in concentration exceeds $10^{20}cm^{-3}$ which is significantly larger than that needed for the charge sheet, therefore preventing its operation. This will mean that the electrons generated in the Germanium region are unable to quickly drift into the high field Silicon region and create an avalanche current, severely

reducing the effectiveness of the SPAD. Another problem seen in this structure is the intermixing of the two bulk materials at the interface (about 50nm). Although not forming a significantly large alloy layer and not being overly detrimental, it should be noted that there is a chance this could create some problems. As absorption at a wavelength of 1.55 μm is at the limit for pure Germanium, any alloying of layers will essentially eliminate the probability of photon absorption. Hence, this alloy layer will effectively reduce the thickness of the absorption region and the overall device efficiency.

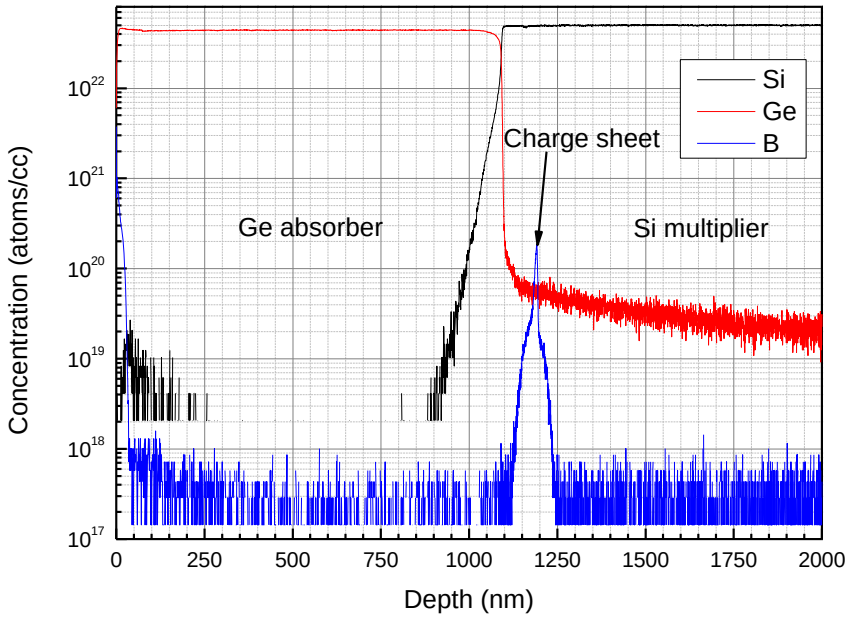


Figure 4.25: SIMS profile of SiGe SPAD1. Doped regions and interdiffusion of the Silicon and Germanium are shown.

Despite this problem, the device could still work, and a thick alloy region would only act as an obstacle for trapping and slowing of the device, if a carrier was generated. A rocking curve of the structure is shown in Figure 4.26. A broadening of the peak suggests a defective Germanium layer, however this is characteristic of Germanium epilayers grown on Silicon. A broad FWHM alone does not constitute a poor layer. Other techniques would be required to confirm this. However, as this layer contains other growth issue which would disrupt the device performance,

more detailed analysis was not performed. Instead, focus was directed towards the optimisation of future structures.

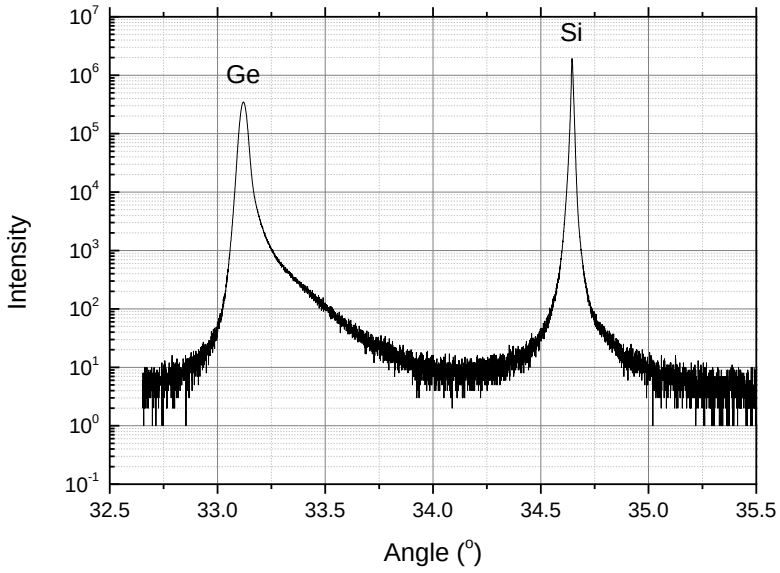


Figure 4.26: (004) rocking curve of SiGe SPAD1. Asymmetry in the Germanium peak is likely to be caused by some intermixing of the two bulk materials at the interface.

Throughout this section several different designs have been structurally characterised. For this work and other work using infrared photon detection the Silicon samples are suitable, but have their place for applications requiring shorter wavelengths. The purely Germanium *pin* structured SPAD also has some potential, but requires thick buffer layers to reduce trapping effects from high TDDs in the intrinsic region. Also as previously mentioned Silicon has advantages over Germanium as the multiplication region. As has been described through the literature, separating the regions for absorption and multiplication provides perhaps the most promising design for SPADs. It is therefore sensible to focus on this design which utilises advantages from each material and can be easily integrated onto Silicon substrates in the electronics industry. The future samples are nearly identical in structure to sample 12-027, but employ a thicker multiplication region.

4.4.5 Charge sheet doping concentration

The doping density in the charge sheet was simulated using Silvaco ATLAS by collaborators, at Herriot Watt University, which is described in further detail elsewhere [88], but was not explored in detail for this work. Three doping densities were used 1×10^{17}, 2×10^{17}, and 5×10^{17} cm^{-3}. With the lowest doping density, described by the blue line in Figure 4.27, the field was too high in the absorber region and could therefore be subject to a higher probability of band-to-band tunnelling. The highest doping density, represented by the black line, meant that punch-through did not occur between the two intrinsic regions and therefore led to inefficient carrier transport. The intermediate doping concentration, shown by the red line, was found to be the best compromise and was therefore selected for these structures [88]. The field had punched through into the absorber region so that photo-generated carriers could be transported to the Silicon layer, but not so high as to create an issue with tunnelling.

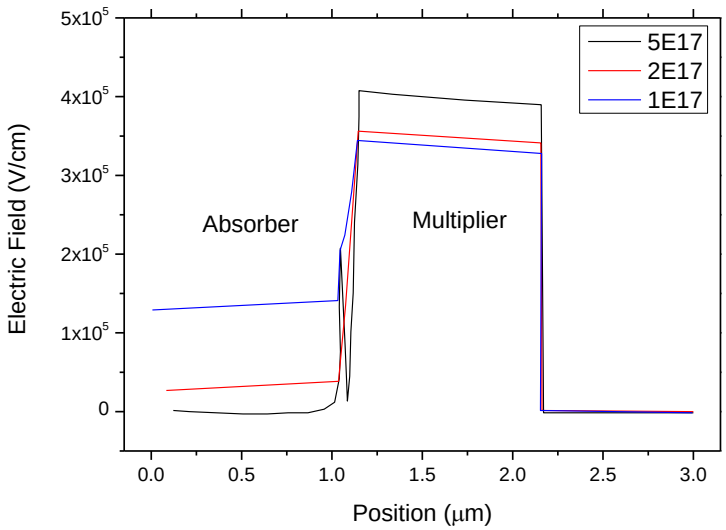


Figure 4.27: Plot showing experimentally determined electric field profile across the absorption and multiplication regions of a SPAD. Adapted from ATLAS data [88].

These simulations were particularly useful in maximizing the chance of fabricating a successful device. The specific doping concentration required highlights the challenge presented for such a device. Growth of this layer must therefore be very accurate if the small window for doping concentration is to be met. This can be difficult to achieve consistently. At these low concentrations there is a relatively

large error, quoted around 50%. This can make calibration difficult, especially when a small deviation is known to be significant for device performance.

4.4.6 New *pipin* Si and Ge SAM SPAD

This section looks at the new *pipin* structured SPAD, and provides detailed structural characterization for the layers. The design, shown in Figure 4.28, follows the separate absorption and multiplication layout, seen in some of the previous structures. The top contact is a heavily p-type (boron) doped Germanium layer. The dopant concentration in this layer is 2-5x$10^{19}$$cm^{-3}$, across the 100nm layer. Below this is the Germanium absorber region, which was chosen to be 1μm thick so that photons at a wavelength around 1330 nm could be absorbed [89]. A much thicker region would increase the probability of absorption, but also increase the transit time into the multiplication region. This layer in nominally undoped, and is generally expected to have a doping concentration of $10^{15}$$cm^{-3}$. The charge sheet layer precedes the absorber region. This is a lightly p-type (boron) doped Silicon layer 100nm in thickness. The function of this layer is to prevent punch-through of the electric field too early, as well as maintain a relatively low (below breakdown) field in the absorber layer. The dopant level in this region is 2x$10^{17}$$cm^{-3}$. Doping this region too heavily would result in the electric field never penetrating the Germanium and therefore preventing efficient transfer of carriers into the high field layer. Doping this layer too lightly would result in a high field region in the Germanium and potential breakdown in this layer. This would render the charge sheet meaningless, as both regions, absorber and multiplier, would experience the high field, as it would be a *pin* diode. The advantage of the separate regions is that the qualities of each material can be fully exploited without being subject to each of their inferiorities.

The first layer grown on the substrate is the Silicon multiplication region. This layer is 1.5μm thick which should provide enough distance for an avalanche current to form, as well as have a suitably low transit time (this is more important in the lower field regions).

Similarly to the Germanium layer the multiplication region is undoped. The bottom contact is the substrate, which is heavily n-type (arsenic) doped. It should be noted that this design was produced before some of the segregation work described

earlier, which is why Arsenic was the dopant used. The dopant concentration was $5x10^{19}cm^{-3}$.

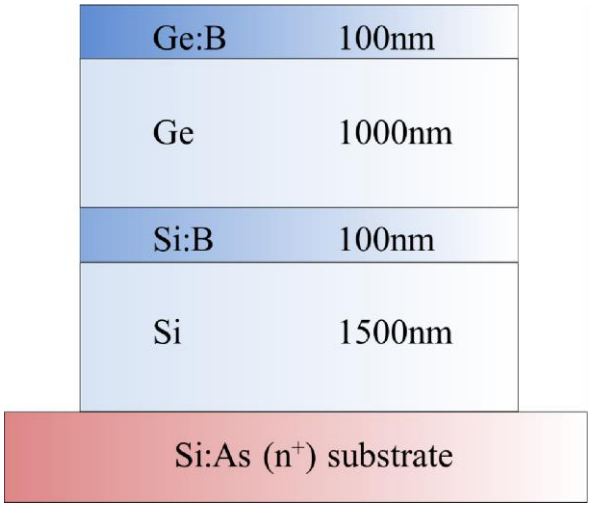


Figure 4.28: Growth plan for pipin SPAD device. The Germanium and Silicon layers are the absorber and multiplier respectively. Sample SiGe SPAD2.

In a device setting the substrate would be the bottom contact, which is produced via etching and passivation, and the top Ge:B layer would act as the top contact. Photons are then incident through this top layer and absorbed in the Germanium region. Carriers are swept into the high field intrinsic Silicon layer where impact ionization occurs leading to an avalanche current, which is collected by the bottom contact.

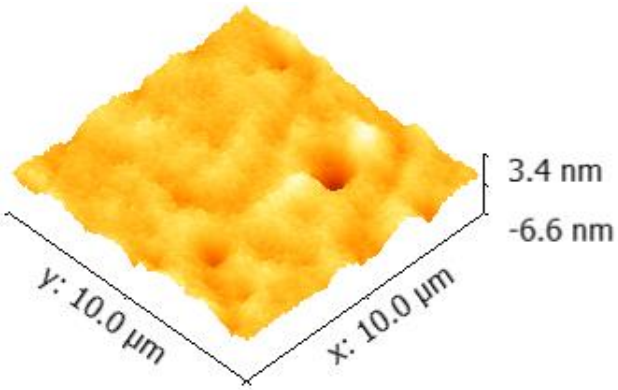


Figure 4.29: AFM image of doped Germanium surface of SPAD sample 13-312, with RMS roughness 0.6nm. Sample SiGe SPAD2.

The smooth topology of these layers, shown in AFM scans demonstrates the successful growth of a low defect Ge absorber region despite the high lattice

mismatch between Si and Ge. The surface of the sample was analysed using AFM (Figure 4.29), where the RMS roughness was measured to be 0.59nm. This is clearly suitable for any device fabrication which is required. This also suggests that there were no significant issues with the growth, which are often characterized by large craters on the surface, where growth has been disrupted by defects deeper in the structure. However TEM imaging must be used to verify this. Several 2D line profiles were taken, which give a representation of the variation in height across the surface. Of all of the AFM scans performed on this sample, the one produced and presented in Figure 4.29 showed the largest variation in height.

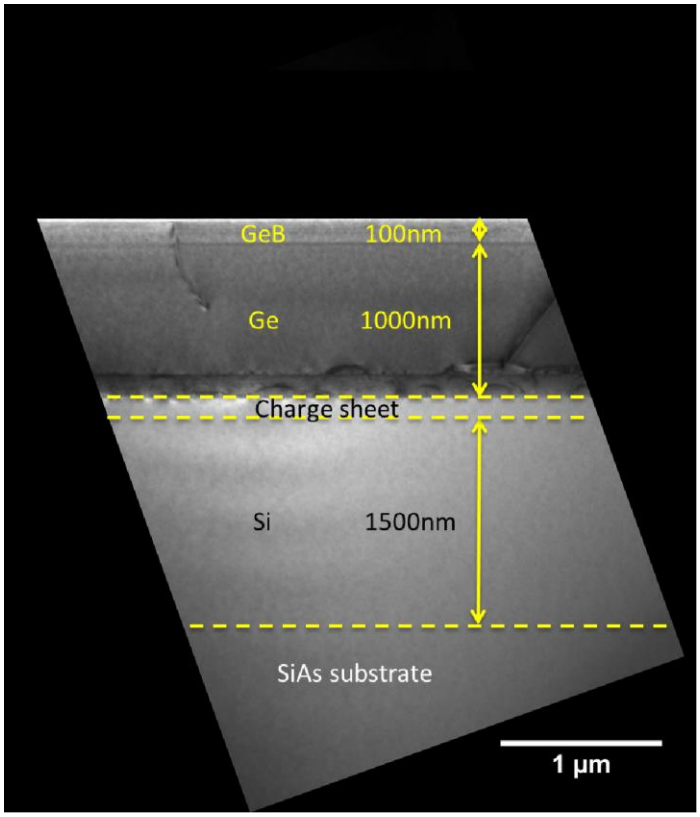


Figure 4.30: TEM image of sample 13-312, taken in (004) dark field diffraction condition. Each separate region can be observed according to the growth plan. Sample SiGe SPAD2.

A TEM image is presented in Figure 4.30, highlighting each region throughout the structure. Most noticeably at the top of the structure is the Ge:B layer. This is mainly distinguished by the growth interrupt marker. It is difficult to see any contrast between this layer and the Germanium layer below due to the doping level being only of the order of 0.1%. Another interesting observation is the clear

multistep growth of the Germanium layer. There is a clear thin region full of dislocations corresponding to the low temperature seed layer. Following this is the much thicker high temperature layer which is evidently much "cleaner" with a much lower TDD. This is very important as the carriers are less likely to fall into a trap and contribute to the afterpulsing effect. Unavoidably, trapping could be an issue while crossing the low temperature Germanium region, but this is clearly a significant improvement on having a thick highly defective layer of Germanium.

Due to the very low doping level in the SiB charge sheet layer (over two orders of magnitude lower than the top contact) no contrast is visible and other techniques are required to make conclusions about this layer. Clean growth from the substrate is observed in the Silicon multiplication layer. As there was no growth interrupt at this stage no line is visible either. A marker has been placed where the substrate should be located, but in the same way as the charge sheet layer, other techniques are necessary to provide evidence of the layer. Overall the TEM imaging has provided useful information about the thickness of the top layers in the structure, as well as confirmation of good crystalline quality throughout the layers. It has also been useful to show that there is only a thin highly defective layer, and therefore verify the two step growth process.

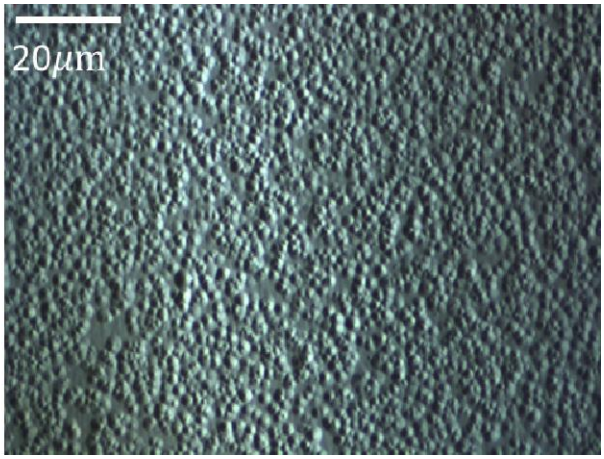


Figure 4.31: HF etched doped Germanium surface of SPAD to demonstrate TDD ($6x10^7 cm^{-2}$) in sample 13-312. Sample SiGe SPAD2.

To corroborate the predictions of low TDD in the Germanium layers made with AFM, and particularly TEM, a chemical etch (described in section 2.8.1) was also used to obtain a value of TDD. An example of one of the images taken is shown in

Figure 4.31. It has been assumed that each pit corresponds to a threading arm penetrating the surface. It should be noted that this assumption may not be entirely true, but as the TDD acquired is an estimate this assumption should be suitable. The value of TDD obtained was around 6 x $10^{7}cm^{-2}$, which is a massive improvement on some other SPAD devices using Germanium which saw a TDD of $10^{10}cm^{-2}$ [45].

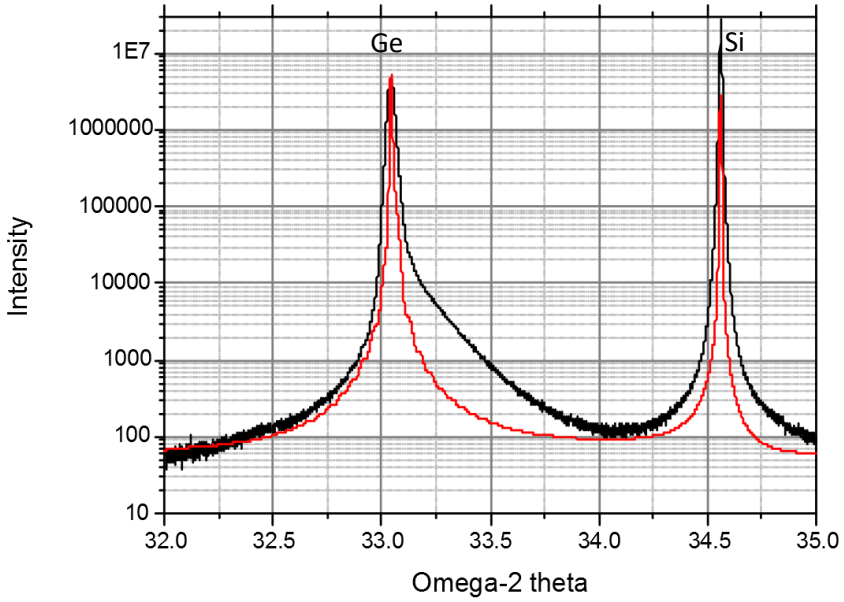


Figure 4.32: (004) rocking curve of sample 13-312, with fitted curve (red). The black line represents the measured data. Sample SiGe SPAD2.

X-ray diffraction (XRD) was used to determine the crystal quality, strain state, and thickness of the layers. Si and Ge peaks are clearly observed in the (004) rocking curve presented in Figure 4.32. Fringes on the Ge peak can be seen corresponding to the Boron doped top contact, which is fully strained to the intrinsic Ge absorber layer. The thickness of this layer was measured to be 106 +/-3 nm, through analysis of these fringes, which is consistent with the XTEM and the original recipe for the structure. The close agreement demonstrates good crystal quality, along with a high quality interface which also indicates that any smearing of the layers has been avoided. The rocking curve shows that the layer is pseudomorphic, further demonstrating the low level of dislocations in the layer (smaller FWHM of peak).

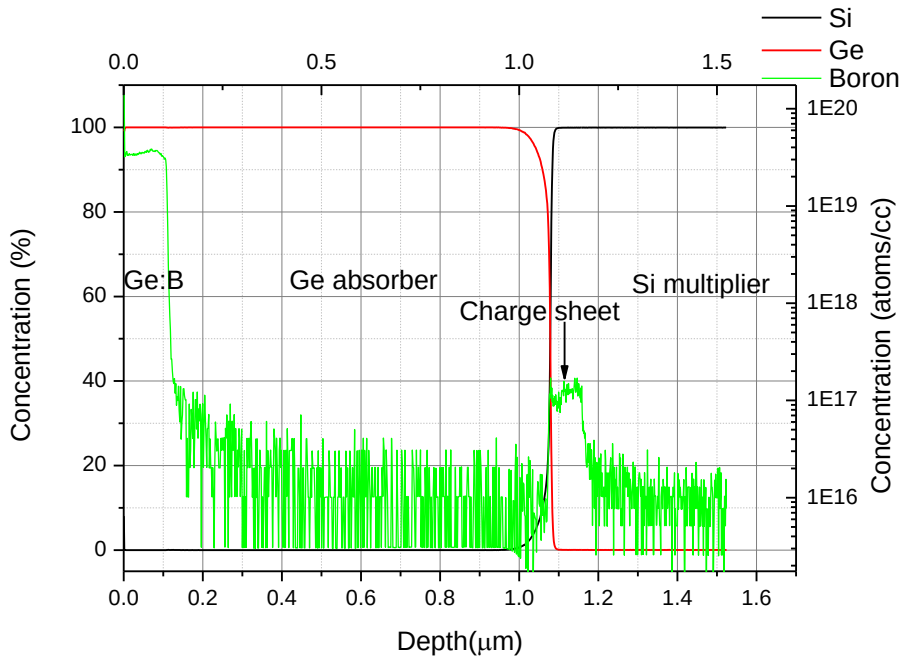


Figure 4.33: SIMS profile for SPAD sample 13-312, including top contact and charge sheet. Sample SiGe SPAD2.

SIMS profiling was used to support the findings in TEM and XRD measurements. The presence of both doped layers can be observed in Figure 4.33. The top contact is represented next to the surface of the sample with a boron concentration of just over $3.5x10^{19}cm^{-3}$. The thickness of this layer, obtained using the FWHM, was found to be 111nm which is consistent with both XRD and TEM. Also the diffusion tail for the Boron into the Germanium layer was calculated to be just 9nm/dec, which confirms that the layer was successfully grown with an abrupt profile. This is important for optimizing the electric field and minimising the presence of acceptors in the intrinsic layer. The dopant concentration in the Germanium region is below the detector limit for the SIMS measurement of around $10^{16}cm^{-3}$ which confirms the intrinsic nature of this layer.

The presence of the lightly doped SiB layer as the charge sheet was not observed in the TEM measurements due to the lack of any contrast between it and the Silicon. However the Boron profile in the SIMS data is clearly evident. The location of the charge sheet is immediately before the start of the two temperature Germanium growth. The boron concentration over this layer is measured at approximately $1.5x10^{17}cm^{-3}$ which should be an appropriate level for the SPAD device [88]. The

abruptness of the doping profile in the charge sheet was measured to be 18 nm/dec which is comparable to that in the top contact. The increase in decay length can be attributed to the diffusion caused by high temperature growth period immediately after the charge layer. Despite this the profile is visibly abrupt which is important for obtaining the required electric field profile.

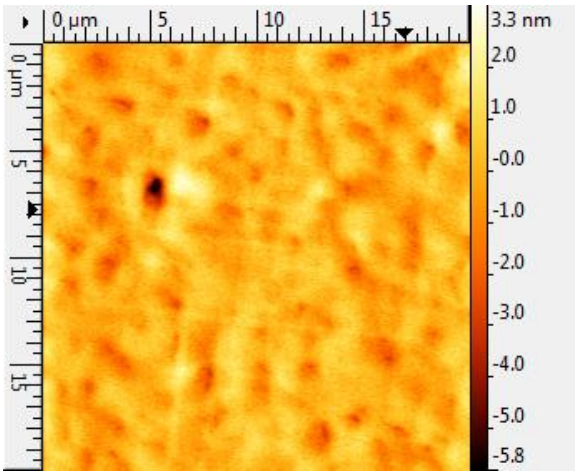


Figure 4.34: 2D AFM scan of Germanium doped surface of SPAD structure. The scale on the right describes vertical height, and the top and left axis describe the horizontal position. RMS roughness 0.59 nm. Sample SiGe SPAD2.

Figures 4.34 and 4.35 demonstrate the reproducibility of the SPAD growth design. The TEM images show a relatively dislocation free Germanium layer, as well as a defined top contact layer, while the AFM scan demonstrates the smooth surfaces of the samples. It can also be seen that there is cross-hatching on the surface, which is caused by the misfit dislocation network and lattice mismatch at the interface between the Silicon and Germanium.

The final part of the characterization for these samples is the bottom contact layer. As previously mentioned, these designs were grown before some of the work on reduction in segregation, which is why Arsenic was used as the n type dopant. Also these structures were grown onto pre-doped substrates.

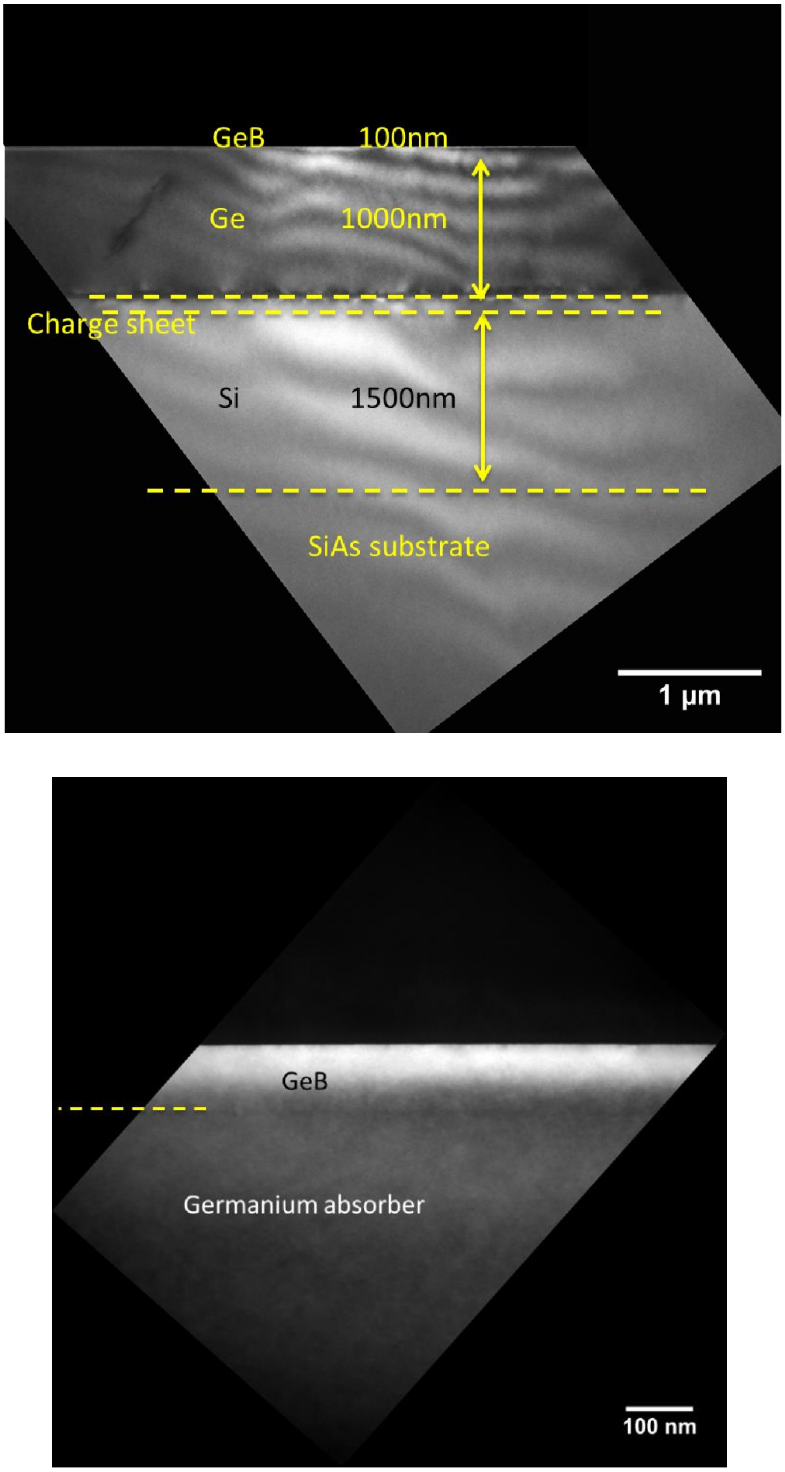


Figure 4.35: Dark field (004) images of other SPAD structures grown to the same design as 13-312. The top image shows the full structure down to the substrate, whereas the bottom image confirms the existence of the highly doped Germanium boron top contact. The white dashed lines represent the doped layers. Sample SiGe SPAD2.

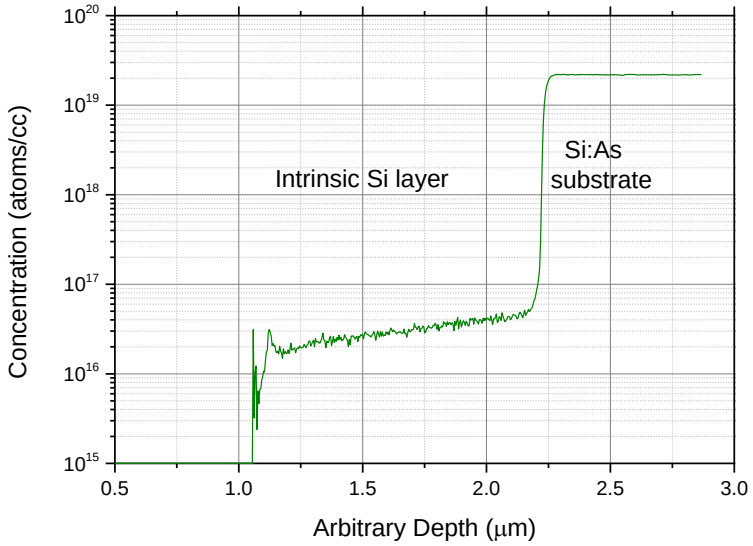


Figure 4.36: SIMS profile of arsenic concentration in Silicon from the ubstrate. The highly doped region represents the doped substrate that the structure was grown on. Sample SiGe SPAD2.

Figure 4.36 shows the concentration of arsenic in the multiplication layer. The concentration of arsenic in this layer was measured to be $2.2 \times 10^{19} cm^{-3}$. This concentration is defined by the substrate used and cannot always be guaranteed. Moreover, the segregation tail was measured to have a decay of more than 3 μm/dec, which produces a concentration of arsenic in the intrinsic layer that approaches two orders of magnitude higher than the ideal design (at its peak). Nevertheless, it is considerably better than the Arsenic doped epilayers from the previous section.

Generally the profiles, crystal quality, and design structure for this SPAD are all suitable for operation as a device. The requirements for each section of the SPAD structure have been successfully met, giving a good chance that this device could make the first single photon detection at 1.55 μm for a Silicon/Germanium design. Also with this current optimization as well as further improvements, the expectation is that these Group IV based designs may become competitive with their III-V counterparts, which is reported later in this work.

There is potential for further optimization of the n-type contact by using epilayers instead of substrate, which give the advantage of easy tailoring of dopant

concentration, as well as the dopant atom used. This is explored in Chapter 5, using the investigation into dopant segregation during further epi-growth.

4.4.7 Replacement of doped substrate with SiP epilayer

The final design of the *pipin* SPAD structure introduces an alternative bottom contact. As was observed in the previous section, the Arsenic doped substrate has an issue with the segregation of the dopant atoms. The introduction of these atoms into the intrinsic layer could potentially disrupt the electric field profile in the device. Unfortunately, as this layer is the substrate, it is difficult to alter the surface concentration or level of segregation of dopant atoms, and therefore the quality of these structures is limited to that of the substrate used.

It has therefore been proposed that an epilayer should be grown onto a lightly doped substrate to act as the bottom n-type contact. This has the advantage that the dopant level may be easily altered if necessary, and the layer can be treated to reduce the amount of migration of dopant atoms from their intended region.

Earlier in this chapter it was found that arsenic doped epilayers showed greater segregation than phosphorus doped layers. When samples consisting of each dopant were grown without any growth tricks the segregation tails for phosphorus were nearly five times steeper than for Arsenic. Although it is speculated that improvements could be made to the Arsenic layers in a similar way to the phosphorus layers, this was not experimentally confirmed. However, due to the superiority of the phosphorus noticed initially it is expected that the phosphorus layers would ultimately produce better results.

The structures in this section follow the growth plan shown in Figure 4.37, where a 300 nm thick phosphorus layer is grown onto a nominally undoped Silicon substrate. During fabrication mesas are etched so that contacts can be made to the bottom layer. 300 nm allows for some error in that etch process.

These doped layers were initially grown at a temperature of 800^{o}C with the subsequent Silicon layer grown at the same temperature. Unfortunately these layers were initially grown just prior to the segregation investigation and at this temperature there is likely to still be a noticeable presence of segregated phosphorus atoms in the intrinsic layer. For the final samples, epilayers would be

used to form the n-type contact consisting of phosphorus as the dopant atom, where the segregation and initial concentration would be reduced using a multi-temperature growth scheme. The low temperature Silicon growth which immediately follows the doped layer would be grown at 500^{o}C and the high temperature layer would follow at 800^{o}C. Depending on the application and if the extra thickness would cause a problem, the multi-step high and low temperature approach could be used instead.

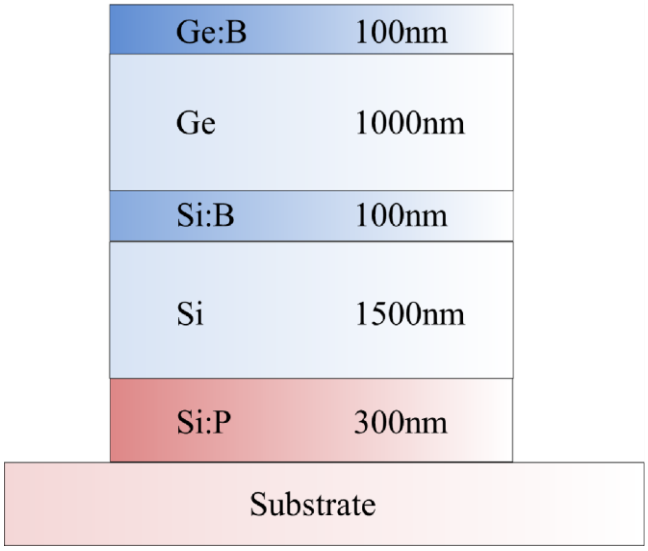


Figure 4.37: SPAD growth structure with epilayer n-type bottom contact. Sample SiGe SPADn.

The doping levels are the same as sample 13-312 for each layer, which are 2-5x10^{19}cm^{-3} for the top contact, 2x10^{17}cm^{-3} for the charge sheet, and 2-5x10^{19}cm^{-3} for the bottom contact. The bulk Germanium and Silicon layers were again, nominally undoped.

As may be expected, the topology of these structures resembled sample 13-312. The surfaces of these samples were very smooth, with an RMS roughness measured to be 0.65nm, and maximum peaks and trough of up to 3 nm (Figure 4.38). The cross-hatching caused by the misfit dislocation network formed at the Silicon and Germanium interface can also be observed. It can therefore be concluded that the introduction of the epilayer n-type bottom contact does not degrade the surface of the structure, which is unsurprising as the concentration of Phosphorus is well below the dopant solubility limit.

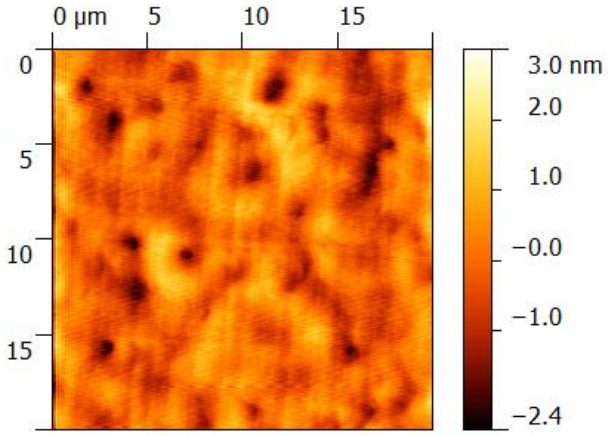


Figure 4.38: Examples of 2D AFM scans for SPAD structure with n-type (GeB) epilayer. Height range is shown on the right axis, while the left and upper axis describe the lateral position. RMS roughness 0.65 nm. Sample SiGe SPADn.

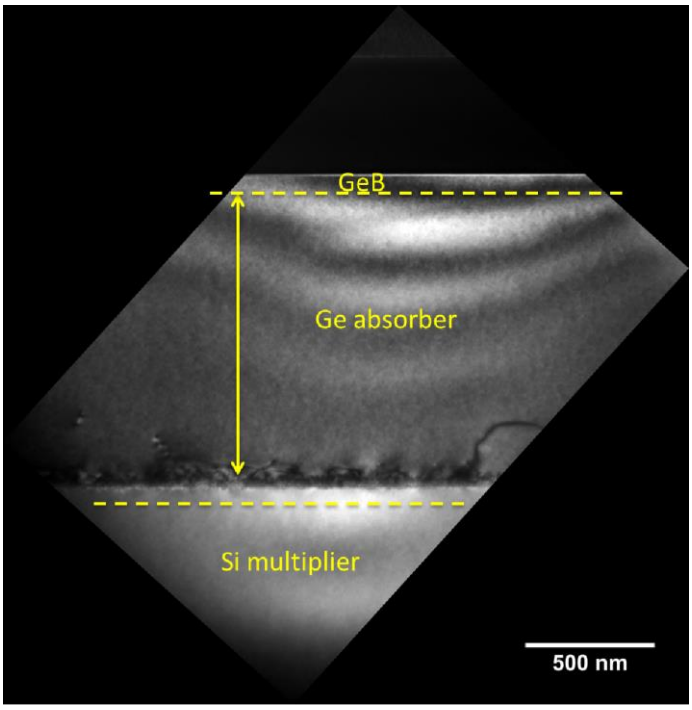
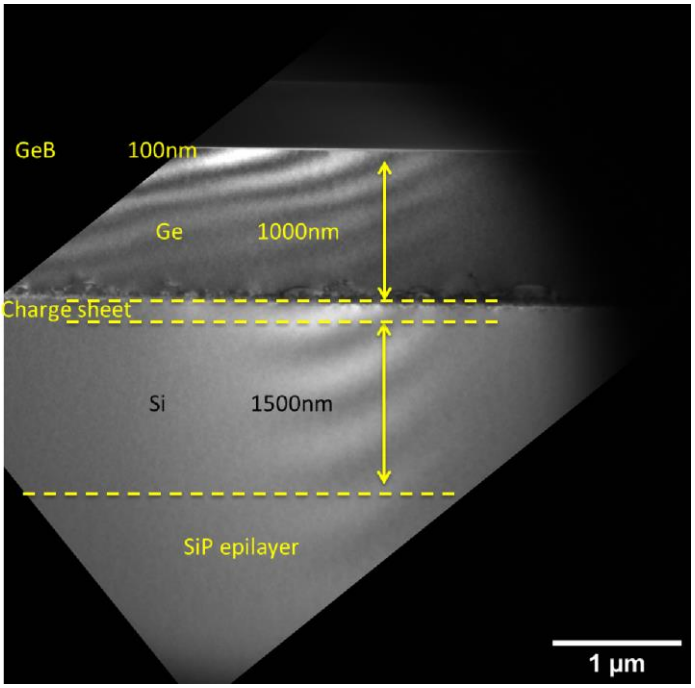


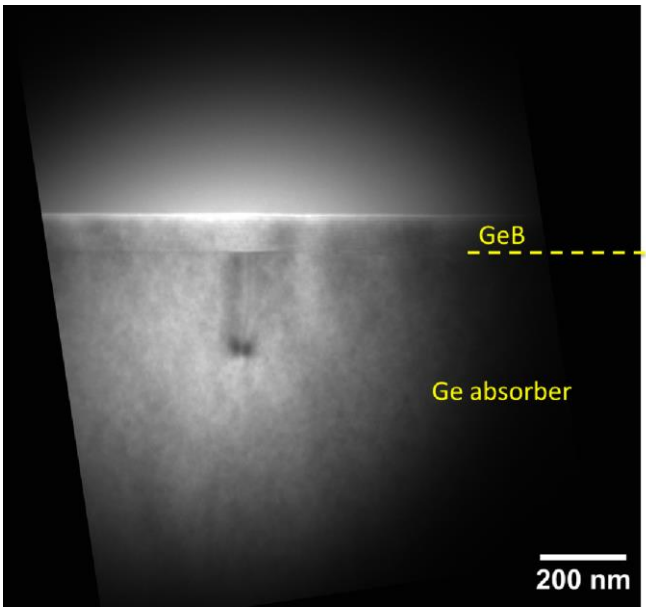
Figure 4.39: Dark field (004) TEM image of Germanium absorber region in SPAD. Sample SiGe SPADn.

To validate that the sample had been grown to the specific growth plan, it was imaged using the TEM. Figure 4.39 shows the absorber region, where it is clear that the dislocation network is confined to the thin region above the interface. This

is beneficial in reducing the region of the device where trapping of photo-generated carriers is high.



(a)



(b)

Figure 4.40 a and b: Dark field (004) TEM image of full structure of SPAD (a) and top p-type (GeB) top contact (b). Sample SiGeSPADn.

The boron doped Germanium top contact can be seen in Figure 4.40a, where the image was taken from a different region of the SPAD structure. The layer is noticed by the growth interrupt, characterized by the line, along with the dislocation terminating at the interface. The thickness of this region is approximately 90nm.

The whole structure is represented in Figure 4.40b, where the dashed lines are used to represent the rough positions of each layer. Due to the low doping concentration in the charge sheet it is not possible to see the contrast in the layer. Also the bottom contact is not visible. This is most likely due to the TEM sample being a bit too thick around the substrate level and not being electron transparent. To test the presence of this layer electrical measurements are required to show the device is acting as a photodetector.

The new SPAD structures were successfully grown to specification, having been characterized structural. Prior to any electrical measurements these structures appear to have a lot of promise. Electrical measurements would be able to identify areas of success, as well as sections which may not be working as intended. Along with the structural characterization performed in this part of the chapter any issues could then be addressed for future improvements of Silicon and Germanium SPADs.

4.5 Electrical characteristics of SPAD device

Dark counts plague the operation of SPAD devices, which is why many operate much more effectively, if not exclusively, at low temperature. A couple of the SPAD devices grown for the purpose of this work have been measured by the author electrically for the level of dark currents.

The measurement of a SPAD is performed by applying a reverse bias across the *pipin* structure. At the breakdown voltage the current should experience a large surge. In an ideal SPAD device the I-V curve for the dark current should follow a similar shape to that of a diode breaking down with increasing reverse bias. Up until the breakdown of the device the current should be very low, however through thermal excitation in the intrinsic region small currents are able to flow.

A drawback of a SPAD is that the device operates more effectively at low temperatures, which can often be impractical. Devices are often liquid nitrogen cooled to 77 K or below, whereas the ideal device could be used at room temperature. The dark currents are clearly much higher when operating at room temperature for example, and therefore are plagued by large DCRs. A possible way to combat this is to only "arm" the device for a short period of time when the arrival of a photon is known, but this is limited to certain applications where this can be anticipated.

Dark current measurements are presented in the section at low temperature and room temperature, along with those when exposed to light (not single photons). Results are also discussed from single photon measurements taken from devices characterized in this work.

Figure 4.41: Optical image of SPAD device. Top contact pad is shown on the right, bottom contact pad on the left. The largest device is 500 µm.

Figure 4.41 shows a fabricated SPAD device where the top contact pad is shown on the right and is connected to the central active area, and the bottom contact pad on the left. The bottom contact pad can be seen to have another square inside the pad which is where contact is made to the n-region. The area in the middle is the active area where incident photons enter the device and can be absorbed in the Germanium absorber region just below the surface of the structure.

These devices were formed by defining cylindrical mesas across the epilayers. A photolithography process was used to etch away the surrounding material all the

way down to the substrate level (bottom contact). For structures which include an epilayer as the back contact, the etching process was targeted to finish at this epilayer. The fabrication of the SPAD devices was performed by Heriot Watt and Glasgow University using fluorine based etch allowed for near vertical side walls. Nickel was used as the metal contact material, owing to its superior low resistivity when compared to other metals used in Silicon and Germanium compounds. The top Nickel contacts were annealed at 325^{o}C, and bond pads were deposited using aluminium.

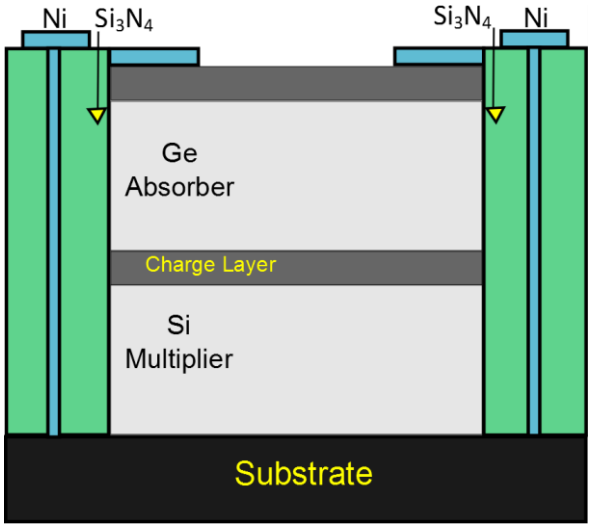


Figure 4.42: SPAD device with fabrication steps, Nickel contacts, and Silicon Nitride passivated side walls.

The side walls were finally passivated with Si_3N_4, and are demonstrated in Figure 4.42. The active area for the mesas was varied between 25 μm and 500 μm.

4.5.1 Dark current measurements for sample 13-191

The first sample which was measured for dark currents was sample 13-191. The structure of this sample was identical to Figure 4.28. TEM images shown in Figure 4.43 indicates that the growth of the structure was mostly as expected except for a roughening at the surface of the sample. This rough surface is also very evident in Figure 4.43, where the RMS roughness was measured to be 6.4 nm.

Room temperature measurements of the dark current were made on several sizes of device, and are shown in Figure 4.44. The breakdown of the devices is characterized by the abrupt increase in the current. The current limit is set by the

probe station used for the measurements, which had a compliance of 40 mA. Five measurements were taken on each device to ensure reliable conclusions could be made. From initial inspection there is a clear variation in device breakdown for the devices which were less than 100μm diameter. However the larger devices show a much more consistent set of curves. There is a steady increase in current over the initial 30V or so which is unlikely to be attributed to the thermally induced carriers in the Silicon. While some of the contribution to this will be those carriers generated thermally, the increase in current is not explained. Some multiplication may occur over the increase in reverse bias, but not as significantly as observed. The most likely reason for this is a fabrication issue which would need to be addressed for future optimization. It should be noted that if punch-through has not occurred in the device thermally induced carriers from the Germanium would not be present in the high-field Silicon layer.

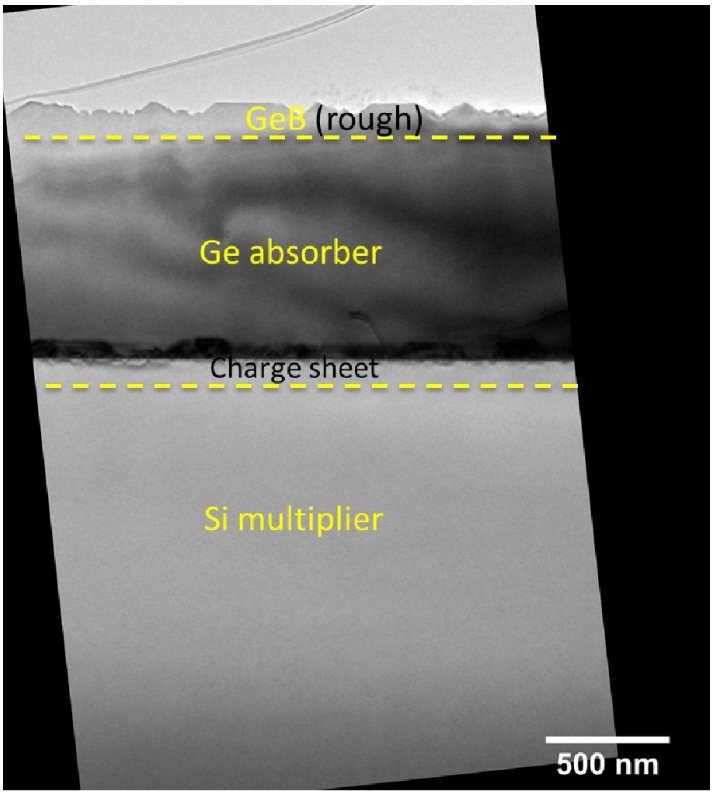


Figure 4.43: TEM image of sample used for fabricated SPAD devices, ID 13-191, taken in straight through condition.

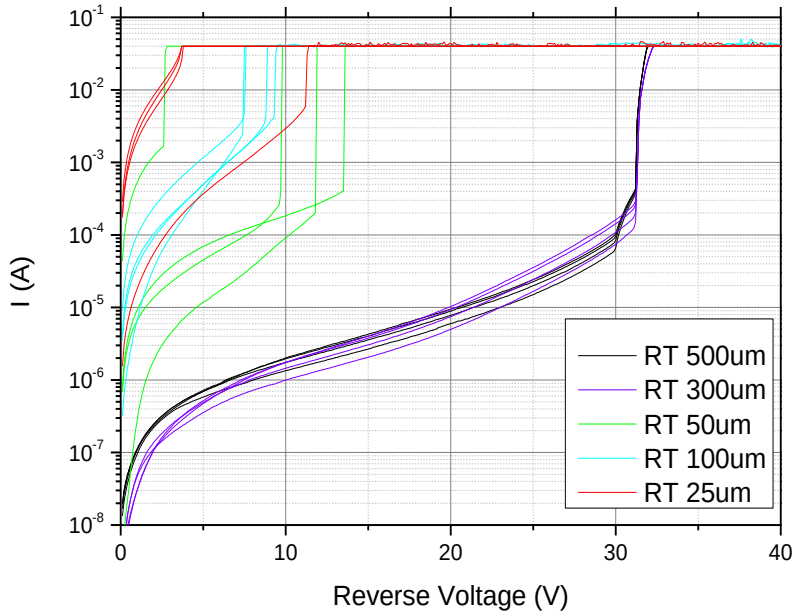


Figure 4.44: Plot of dark currents for different mesa sized (diameter) SPAD devices (13-191) measured at room temperature. Reverse bias is applied, using a probe station, to each SPAD device and is increased up to 40V (instrumental limit). Breakdown of each device is observed by the sudden increase in current up to the saturation limit (0.04A).

For the two largest devices breakdown occurs at 31.3 V (at RT). There is a sudden rise in current which is caused by avalanche breakdown, which is expected for SPAD operation. While the smaller devices do experience a thermally induced current prior to a breakdown region, the bias at which these events take place is vastly different from measurement to measurement. One proposed conclusion is that the device is breaking down along the side walls due to an issue with the fabrication, which is more significant for the smaller devices where the surface:volume ratio is greater.

The same selection of devices was also measured at low temperature (77 K), where a similar trend is observed to that seen at high temperature (Figure 4.45). The smaller devices displayed varied currents and breakdown voltages, whereas the larger devices were much more regular in their characteristics. There are two important observations for the larger devices at low temperature: the first is the much reduced current while at voltages below breakdown; the second is the shift in breakdown voltage. As expected at the lower temperature there are far fewer free

carriers available through thermal generation, which is evident in the low current recorded (up to 4/5 orders of magnitude different). The steady increase in dark current for the smaller diameter devices is still present which further demonstrates that its nature is due to fabrication problems. The shift in the breakdown voltage is also expected. At the lower temperatures there are fewer thermal vibrations of the bulk atoms. This allows for free carriers to accelerate further without colliding with a phonon, which means they are able to reach higher energies (suitable for impact ionization) at lower voltages than at room temperature (when they are likely to lose energy more frequently from phonon scattering). The breakdown for the larger devices is seen at a bias of 27.8V (at 77K).

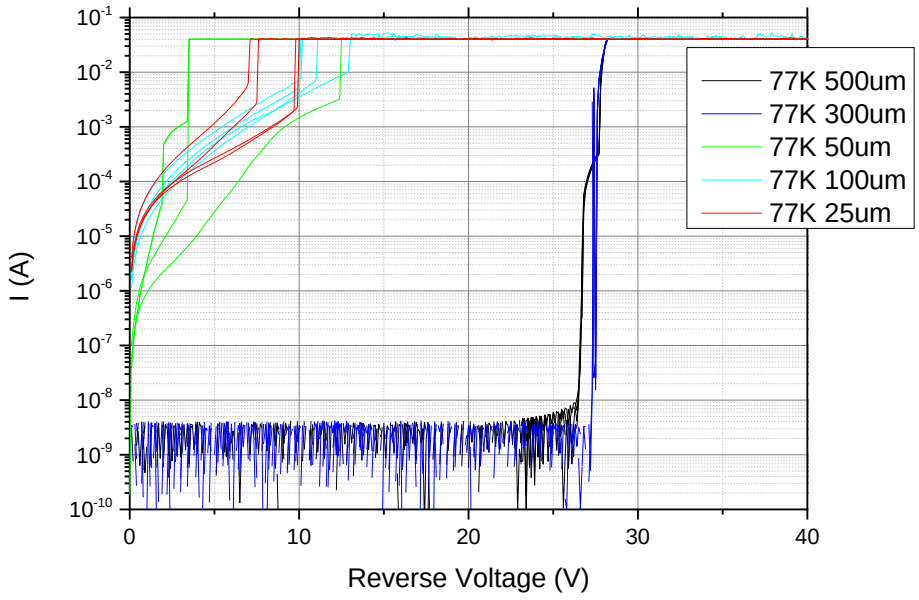


Figure 4.45: Plot of dark currents for different mesa sized SPAD devices (13-191) measured at low temperature (77 K). Reverse bias is applied, using a probe station, to each SPAD device and is increased up to 40 V (instrumental limit). Breakdown of each device is observed by the sudden increase in current up to the saturation limit (0.04 A).

It can be concluded that these devices are much more suitable for operation at lower temperatures, which is true for all SPAD devices. Therefore with suitable gating of Geiger mode operation the chance of a current caused by the arrival of a photon, as opposed to that of a dark event, is more likely. Ideally these devices would be used at room temperature, but it can clearly be seen that operation would

be plagued by dark currents, and much shorter gating of the SPADs would be required, along with known arrival times of the photon.

4.5.2 Dark current measurements for sample 14-329

A second batch of the newest SPAD devices were fabricated (on sample 14-329) by Herriot Watt and Glasgow University before in-house dark current analysis was performed. The dark measurements gave a useful indication of how effective normal operation may be and how much the dark currents may affect detection. Light measurements were performed under non single photon conditions at Warwick University to ensure the devices were capable of detecting light.

Dark current measurements were performed on sample 14-329, which followed the newest growth plan, containing an epilayer for the n-type bottom contact. Several features can be noted from Figure 4.46 for this device. Firstly, the device does not breakdown before 40V reverse bias is applied. Unfortunately, the parameter analyser used to perform the measurements is limited to 40 V. Secondly, there is a charging of the device, Figure 4.46 highlights this problem. The black line represents the first measurement which was taken, and the red and blue lines are second and third respectively. As can be observed the current through the device decreases during each run, until a saturation point. Measurements were taken until no significant change was noticed which occurred after 5 repeats. It would appear that after the initial measurement a significant number of carriers remain at the contacts or traps along the passivated walls, which results in a lower current for subsequent runs. The lowering of the dark current is actually beneficial; however, this does suggest a potential issue with either the structure or fabrication procedure.

From an application point of view this would affect device operation, as the SPAD would be operating as though it was “charged”, therefore follow the blue dark current curve from Figure 4.46. Assuming that the dead time was relatively short, which is often the situation for SPADs, the device would not have time to discharge and return to the base state. In the first instance the device should be broken down to obtain the charged state. A crude test showed that the delay on the devices for them to return to the original state was in excess of 3 hours. Therefore it could be assumed that the normal state for the dark current is much closer to that of the saturated curves on the IVs. It should also be noted that the equipment was

grounded in between measurements to ensure that the issue was not caused by the equipment.

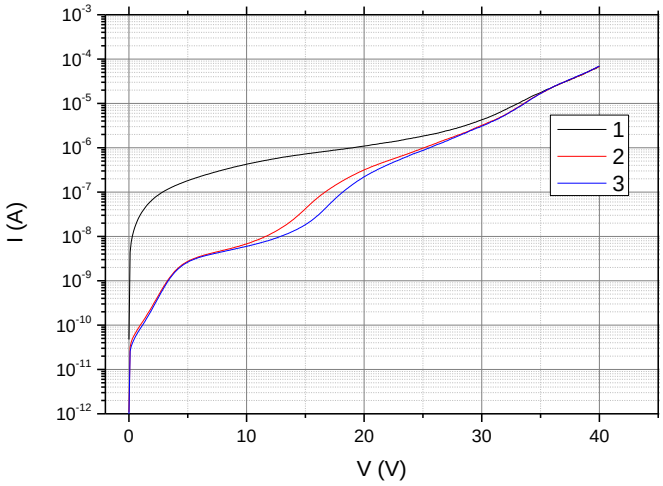


Figure 4.46: Plot of dark current for SPAD device (14-329) measured at room temperature. Reverse bias is applied, using a probe station, to each SPAD device and is increased up to 40 V (instrumental limit). 1, 2, and 3 represent the 1st, 2nd, and 3rd repeat of the measurement.

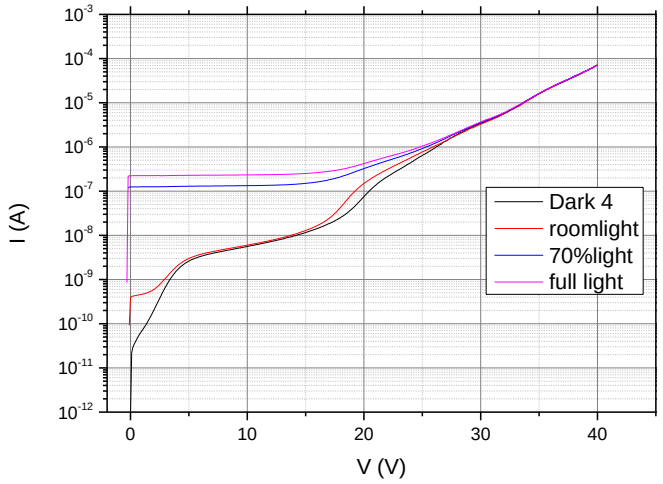


Figure 4.47: Plot of dark current for SPAD device (14-329) measured at room temperature. 4 different "light levels" were used to test the devices photoresponsivity. Reverse bias is applied, using a variable temperature probe station, to each SPAD device and is increased up to 40 V (instrumental limit).

To check that the devices responded to the presence of photons, light was incident on the samples for a number of measurements. This was performed by simply removing the cover on the probe station which blocked out light, and by using the camera light as the source of photons. This was not indicative of single photon detection, nor was a quantifiable number of photons measured, but instead acted as a useful way of testing general photon detection. The curves displayed in Figure 4.47 confirms the ability of the devices to detect photons. There is a small photo induced current when the sample is subject to only ambient light from the room, while there is a significant increase in the current when the lamp is illuminating the sample.

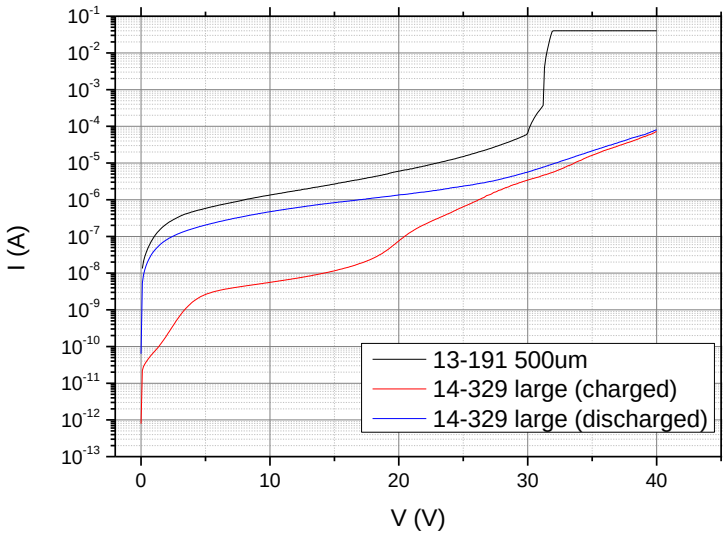


Figure 4.48: Comparison of room temperature dark current measurements with increasing applied reverse bias for samples 14-329, and 13-191.

A comparison is made in Figure 4.48 to sample 13-191. The main difference is that there is a change in the breakdown voltage, which is most likely due to a difference in the thickness of the multiplication regions. When focusing on the dark currents measured the newer sample (14-329) displays an improvement to the reduction of the dark current, even when compared to the base state. When given time to discharge fully there is up to an order of magnitude difference in the two curves, and for the charged state over two orders of magnitude. This is likely due to an improvement in structure quality, although it could be a result of other factors such as an improved interface quality, which could have been poorer when using the

substrate as the bottom n-type contact. Alternatively the dislocation density could be lower around the Germanium-Silicon interface, leading to fewer trapped carriers. Also fabrication improvements, for example side wall passivation, could also reduce the leakage paths for carriers and therefore dark current.

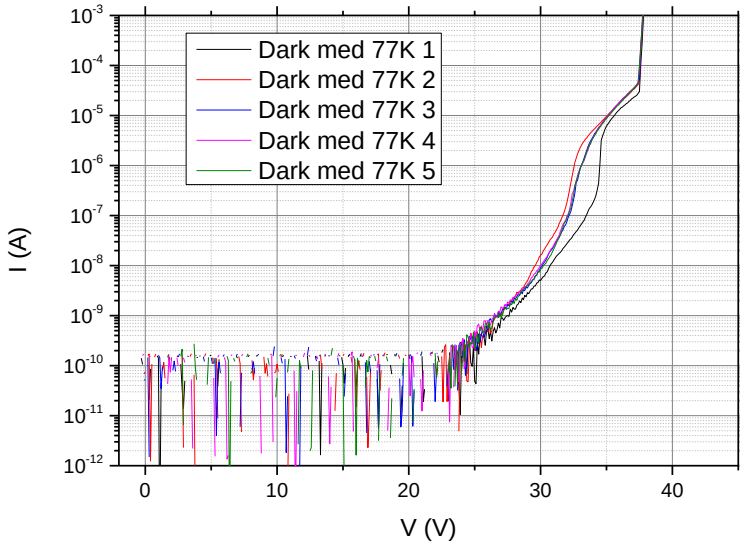


Figure 4.49: Plot of dark current for SPAD device (14-329) measured at low temperature (77K). Reverse bias is applied, using a probe station, to each SPAD device and is increased up to 40V (instrumental limit). Breakdown is observed at the point of sudden increase in current.

Sample 14-329 was cooled down to 77 K to investigate the dark current at lower temperature. Predictably the current, shown in Figure 4.49, is much lower at 77 K, due to the lack of thermally generated carriers in the structure. There is a shift in the breakdown of the device at low temperature, which is expected due to the increased distance between scattering events leading to higher kinetic energies obtained causing impact ionization. Breakdown was observed at a reverse bias of approximately 33 V.

Also shown here the charging effect, although present, was noticeably less pronounced than at room temperature. This could be another consequence of reduced scattering at low temperatures, allowing the carriers to move more freely and the device to discharge. Therefore the charging effect seen in this device is not

so problematic. It is thought that the charging could be a sign for further optimization of the fabrication process.

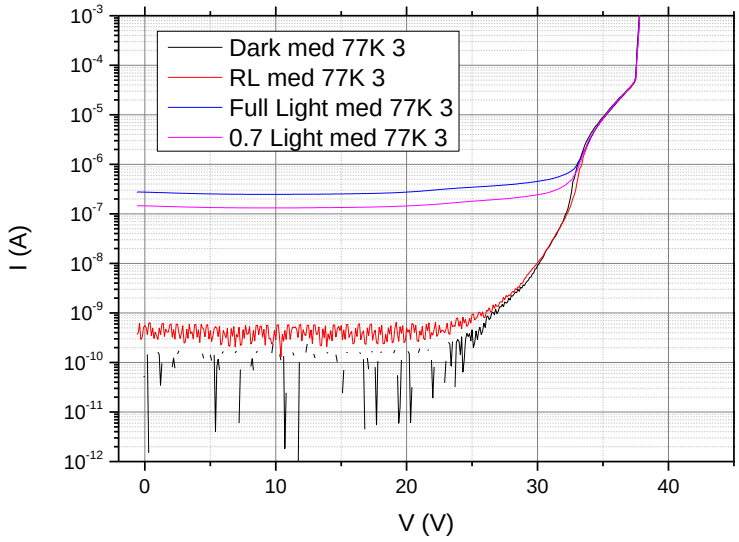


Figure 4.50: Plot of dark current for SPAD device (14-329) measured at low temperature (77 K). 4 different "light levels" were used to test the devices photoresponsivity. Reverse bias is applied, using a variable temperature probe station, to each SPAD device and is increased up to 40 V (instrumental limit).

Similarly to the room temperature measurements, the sample was subject to light to prove photo detection capability. The increase in difference between light and dark currents in Figure 4.50 when compared to room temperature measurements is only caused by a lower dark current. The current when exposed to light is essentially the same as the photo generated current as it outweighs the dark counts significantly. The increase in dark current with increasing reverse bias before breakdown is still evident for these structures. However there is a decrease for device 14-329, which suggests some improvement to fabrication, aside from the charging effect. The base current is also slightly lower, observed in Figure 4.49, which may show a reduction to thermally generated carriers, which is possibly due to improved structure quality.

Figure 4.51 shows the comparison between the room temperature and 77 K results. Despite not breaking down the current around 40 V, for the room temperature curves, is approaching mA, due to the significant number of thermally generated carriers and subsequent secondary carriers. The slight bumps in the curves could be

caused by parallel conduction through the passivated side walls. It is possible for carriers to conduct through the side wall as this is an “easier” path to take. This would indicate that there are further improvements to this area which could be made so that the dark current could be reduced further. Although very difficult, if there were no conduction paths through the sides, the device would be more effective under single photon operation, as dark currents from all sources are a huge drawback. Equally any non-optimized growth issues are also likely to lead to problems, and possibly device failure.

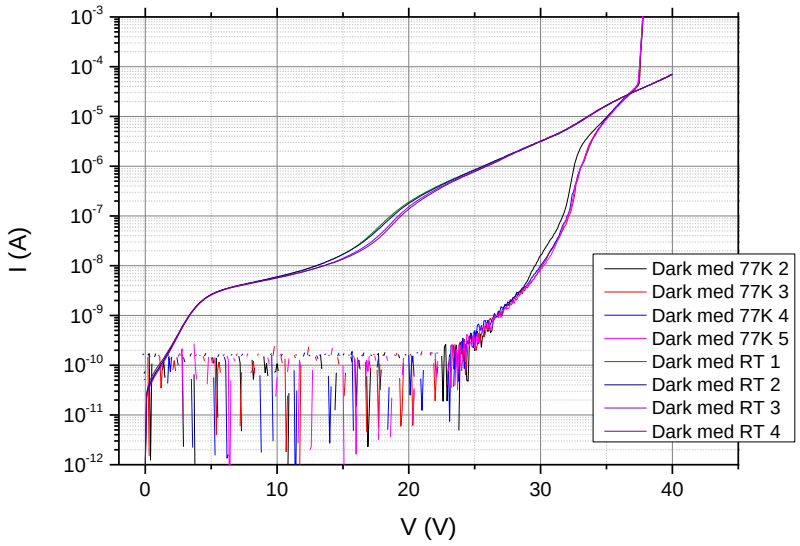


Figure 4.51: Plot comparing dark current for SPAD device (14-329) measured at low temperature (77 K) and room temperature. Reverse bias is applied, using a probe station, to each SPAD device and is increased up to 40 V (instrumental limit). Breakdown is observed at the point of sudden increase in current.

4.6 Single photon detection for SPAD sample 13-312

Fabricated devices using the successful structures (13-312) described in this chapter have been experimentally tested for single photon detection. The structures used had the substrate as the n-type contact. The precise details of the set up and electrical characterization of these devices can be found in the journal [88]. As a general description, the samples were mounted into a cryostat before being illuminated. The photon flux was less than 0.1 photons per pulse, which is important in ensuring that detection is of single photons [88]. This is obtained using weak coherent pulses, which

are defined as laser pulses containing less than 1 photon per pulse. This value can be tuned by attenuating the laser pulse accordingly. Using poissonian statistics is can be shown that that at 0.1 photons per pulse, very few (around 5%) pulses will contain more than one photon.

Measurements were performed in gated mode, where the biasing of the device was normally kept a few volts below the breakdown voltage, and the pulses where the device entered Geiger mode were around 10ns. After an avalanche, active quenching was used to reset the SPAD.

The devices were tested at 1330 nm and 1550 nm. Testing revealed single photon detection efficiencies comparable to that of their III-V counterparts at 1330 nm, as well as the first reported detection of single photons at 1550 nm. Naturally the detection efficiencies at the longer wavelength were lower than that at 1330 nm due to Germanium's lower absorption coefficient at this wavelength. The plots of these results are presented in Figure 4.53.

The DCR was also measured to be between 10^6 and 10^7 counts per second. An advantage of this was that there was a significantly low portion of this which was caused through afterpulsing, particularly when compared to the III-V devices. High gating frequencies were able to be used with relatively low afterpulsing effects. This is very beneficial in a practical sense as the device can be reset quickly, therefore spending less time in a "dead" state, which is common in devices made from InGaAs/InP.

It was predicted that with further optimization of the fabrication process through side wall passivation improvements to the DCR could be dramatically lowered. Also it is hoped that with the further improvements to the structure, including reduction of segregation into the intrinsic region device characteristics could be improved yet more. Much of the structural optimization consists of accuracy in the doping and reduction of atom migration. Attempting to improve this has been appreciably difficult and it is apparent that small deviations from the ideal structure can cause the device to be unsuccessful.

There are many factors which ultimately determine the effectiveness of a SPAD device, both from a growth side and a fabrication side, where a perfect device would

be fully optimized in both areas. This is a serious challenge, and work in this section has provided some optimization techniques to allow for a more successful device. It has culminated so far in the first detection of single photons at a wavelength of 1550 nm, while improving that of detection at 1330 nm, both of which have application in the communication at infrared wavelengths. Also steps have been made to compete with the detection capabilities of III-V SPADs, which would be of great benefit because of the Group IVs compatibility with many other Silicon based technologies all on one platform.

Initial measurements were carried out at Herriot Watt University to test for non-single photon detection at 1550 nm. Although these are just preliminary results which confirm the capability of the devices to detect at such wavelengths, they are crucial for checking that the device could, in principle, detect a single photon of 1550 nm wavelength.

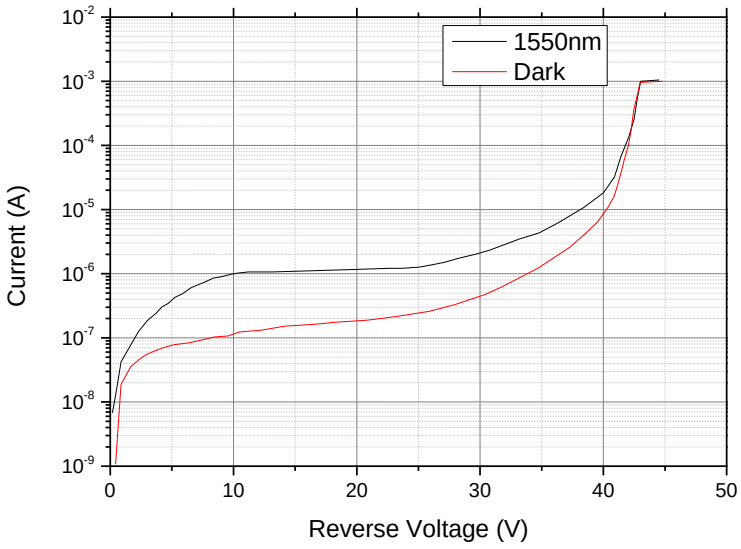


Figure 4.52: Initial results demonstrating photon detection at 1550 nm and under dark conditions for new structure SPAD device (14-329) (100 μm diameter active area).

The dark current measurement (from Herriot Watt) has been included to show the difference between those carriers generated thermally and those by photons. There is a slight increase associated with the increase in reverse bias for the dark current as seen for previously analysed sample (13-191). Again this is likely to be due to a

fabrication issue. Importantly there is clear evidence (Figure 4.52) of sensitivity to the infrared wavelength, but as no single photon measurements have been performed on the batch of samples containing sample 14-329 to date, it is not yet possible to confirm the expected improvements.

A device was fabricated using the structure (13-312) designed throughout this chapter. This was tested at Herriot Watt University for its single photon detecting capabilities at 1310 nm.

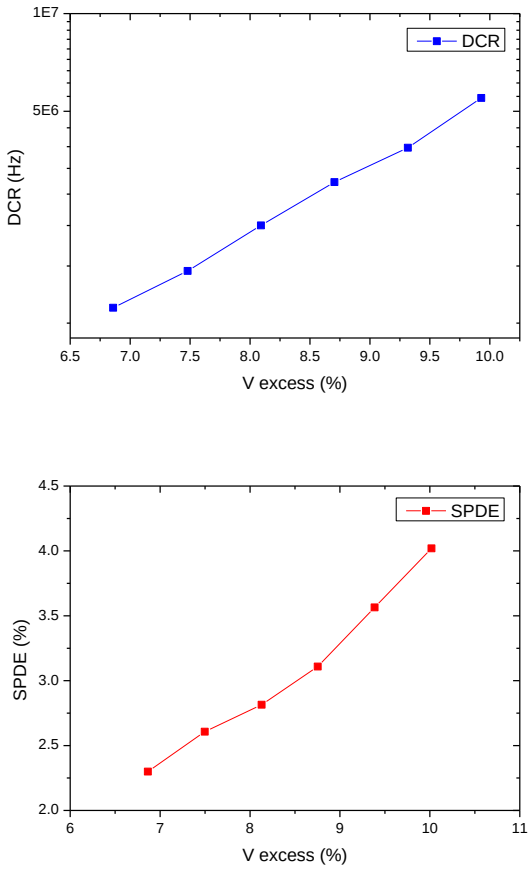


Figure 4.53: Single photon detection efficiency (SPDE) at 1310 nm, and DCR for SPAD sample (25μm diameter), as a function of excess bias for sample 13-312.

There are two main results to be taken from the results presented in Figure 4.53. Firstly, the peak detection efficiency of 4% has brought the Ge-on-Si SPAD into a position of being comparable to InGaAs/InP SPADs. Secondly, the SPDE increases with excess bias. While the actual absorption of the photon is not dependent on the

bias voltage, a greater reverse excess bias (where excess bias is the bias above breakdown) will result in an increased probability of initiating an avalanche. Although not presented in this graph, the detection efficiency at 1550 nm was measured to be around a factor of 10 lower than that at 1330 nm. This is to be expected due to the decreased likelihood of photon absorption in Germanium for this wavelength. However, this is nevertheless a significant result, as it is a first for a Germanium on Silicon device measured at 1550 nm.

The dark count rate of the device is comparable to the III-V devices, but the device is superior with respect to afterpulsing effects. A maximum DCR peaked at $5x10^6$Hz was measured, however this was at a large excess bias of 9.5% Likewise with the DCR, an increased excess bias results in a higher number of false counts. This introduces a need to find an optimal operating regime for the device, as an increase in bias can improve the overall SPDE, but will also increase the DCR.

4.7 SPAD Summary

The key electrical measurements are the detection efficiency and the dark count rate for a SPAD device. Below in Table 4 a summary is presented for different SPAD devices designed with a selection of different materials over a range of wavelengths. The dark count column does not take into consideration the active area size so not all samples are directly comparable.

To show how the device compares to the current state-of-the-art Figure 4.54 has been plotted. The devices grown in this work are circled in blue, and the other devices are numbered correspondingly. As can be seen less work has been done on devices at longer wavelengths approaching the infrared and very little has been performed on devices made from material other than InGaAs/InP. The attempts on group IV materials have been minimal and largely unsuccessful. While one device, a planar Germanium SPAD [46], has shown detection efficiencies higher than any other at 1550 nm, the Germanium on Silicon SAMSPAD is the first of its type to detect photons at this wavelength. It also shows a detection efficiency comparable to the III-V devices, and has exceeded some of those devices in that category.

Table 4: Summary table of a selection of SPAD devices, from literature, using different materials at different wavelengths. Detection efficiency and dark count rate are reported.

Ref	**Material**	**Author (ref)**	**DE(wavelength)**		**DCR (cps)**
1	SOI	Lee et al. [90]	25.4%	(490nm)	2.76×10^4
1	SOI	Lee et al. [90]	7.7%	(850nm)	
2	Poly Si	Niclass et al. [91]	41%	(450nm)	6.70×10^5
2	Poly Si	Niclass et al. [91]	3.8%	(850nm)	
3	Si	Richardson et al [92]	28%	(500nm)	
4	Si	Louden et al	0.0001%	(1210nm)	1.00×10^7
4	SiGe	Louden et al	0.005%	(1210nm)	1.00×10^7
4	Si	Louden et al	1%	(826nm)	
4	SiGe	Louden et al	0.5%	(826nm)	
5	Si	SensL	20%	(500nm)	
5	Si	SensL	6%	(800nm)	
6	InGaAs/InP	Lacaita et al [93]	0.04%	(1300nm)	
6	InGaAs/InP	Lacaita et al [93]	0.1%	(1300nm)	
7	InGaAs/InP	Hiskett et al.	15%	(1550 nm)	3.00×10^5
8	InGaAs/InP	Pellegrini	10%	(1550 nm)	$1x10^2$ - $1x10^6$
9	InGaAs/InP	Tosi	20%	(1550 nm)	5.00×10^5
10	**Ge on Si**	**This Work**	**4%**	**(1310nm)**	**1.00×10^6**
10	**Ge on Si**	**This Work**	**0.4%**	**(1550 nm)**	
11	Ge	Carroll et al.	0.000001%	(1550 nm)	2.80×10^5
12	Ge Planar	Tosi et al.	1%	(1550 nm)	1.00×10^7
13	InGaAs/InP	Liu et al [94]	45%	(1310nm)	1.20×10^4

The main conclusion to this is the vast improvement in detection at the longer wavelengths when compared the group IV predecessors. This is particularly notable because of the low efficiencies seen for Silicon and the Silicon Germanium alloy devices at wavelengths beyond 1 μm.

While there is still a gap between the two different materials, InGaAs/InP and SiGe, the comparable efficiencies and dark counts, and the advantageous afterpulsing effects, make the Group IV devices an appealing alternative to the III-Vs. Especially when there is still optimization to be reached.

A useful statistic for SPAD devices is the noise equivalent power (NEP), which incorporates the detection efficiency and DCR of a device. This is calculated through $NEP = \frac{hc}{\lambda\eta}\sqrt{2N_D}$, where the quantum efficiency and dark count rates are η and N_D respectively. For the device which was tested under single photon conditions the NEP was $1x10^{-14}WHz^{-1/2}$. It is thought that with further optimization, particularly around the sidewall passivation, which is known to give rise to dark counts through leakages, this value could be further reduced. It has been shown that there is a correlation between the TDD, which is prevalent in Ge on SI structures, and the leakage current (dark current factor). Therefore along with fabrication improvements the reduction in TDD would also be advantageous in reducing the NEP of the device [49]. Previous Silicon and Silicon-Germanium devices in a vertical structure have recorded NEPs of $1x10^{-11}$ and $5x10^{-12}$ $WHz^{-1/2}$ respectively [59]. III-V devices tend to have lower NEPs of around $10^{-16}WHz^{-1/2}$ [55].

The InGaAs devices tend to have lower DCRs at the lower temperatures. While this device showed improvements to other Ge APD/SPAD devices, there is still some improvement required to match the III-Vs. This is thought to be enhanced with fabrication. One advantage which is also worth mentioning is that the DCR rises significantly with temperature for the InGaAs SPADs while the Germanium on Silicon device shows a less severe increase (likely due to their decreased afterpulsing effect). Most of the leading III-V device results use a repetition rate (dependent on pulse time and dead time) of around 10kHz, whereas even the first SiGe device was capable of this. This corresponds to a dead time of around 100μs with pulses of the order 10ns. The Germanium on Silicon device characterized in

this work was tested up to a repetition rate of 1MHz (approx. dead time of 1μs) without a severe change to DCR. This is thought to be because of the superior afterpulsing effects when compared to III-Vs. Therefore with such a quick turnaround possible, the device could be more efficient.

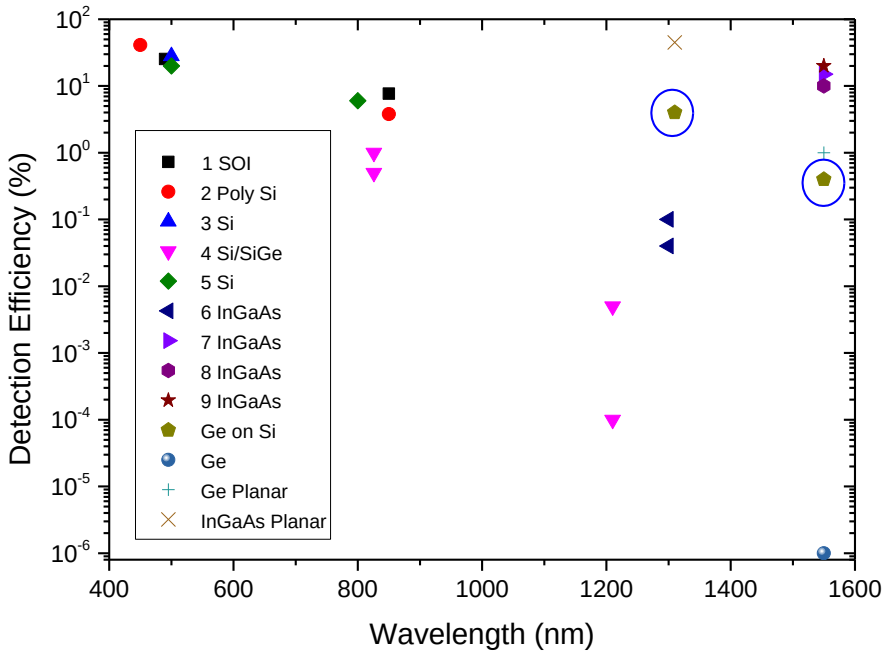


Figure 4.54: Summary graph of a selection of SPAD devices using different materials at different wavelengths. Data points are taken from Table 4, with SPDE results from this work also included (circled).

4.8 Suspended SPAD device concept

With the introduction of an epilayer n-type contact, the substrate is rendered essentially redundant after the growth process. An interesting option for future SPAD designs is that of a suspended device. Potentially there could be unwanted conduction through the substrate which could affect the performance of the device. However, the main attraction of removing the substrate is the opportunity to then introduce a reflective layer directly under the device. This could reflect any photons which are not absorbed in the Germanium layer on their first pass through the device. The reflected photons then have the chance of being absorbed on the

second pass through the layers which would potentially double the overall detection efficiency [43].

Introduction of a metal reflective layer is difficult. If this was deposited prior to the SPAD growth the mismatch and difference in atomic structure would result in very poor growth quality. Upon removal of the substrate the back contact is exposed. It would then be possible to evaporate the reflective metal onto the contact while maintaining a perfectly grown structure.

The potential problem with this is the etch resistance of Si:P. If it is not resistant it cannot act as an etch stop and the etching process becomes impossibly difficult. An alternative option would be to grow an etch resistant layer before the bottom contact layer which is transparent to infrared light.

Preliminary work required to explore this suspended SPAD device concept forms the content of the next two Chapters.

Chapter 5

Very Low Resistance Ohmic Contacts for SPAD Devices

5.1 Motivation for highly doped SiB layers

The doping of group IV semiconductor materials has long been of interest, as the properties of materials such as Silicon and Germanium can be altered with a relatively small number of foreign atoms. If a Ge on Si SPAD device was to be realised on SOI instead of on a normal Silicon substrate, a highly doped bottom contact layer would be essential. This would require optimal growth of such a layer, since the conducting substrate could no longer provide the bottom contact for the SPAD device.

P and As are the most common n-type dopants, which would be useful for the *pipin* structure bottom contact region. Antimony is another candidate for doping, but is not considered in this work because it is incredibly difficult for it to be grown using CVD, partly due to the difficulty in incorporating such a large atom [95]. Prior to this work, P and As doping concentrations were unable to exceed $4x10^{19}$ cm^{-3}, whereas more potential had been shown for using boron as a p-type dopant source instead. Despite the fact that this is already a high doping concentration, for applications with regard to etch resistance, it is unclear whether or not this is high enough. It was therefore decided that this work presented in the following chapter would focus on the p-type dopant. It should be noted that if this layer was used for a bottom contact for a SPAD device the structure may have to be reversed to be *ninip*. With regards to a top contact layer, a SPAD device requires a highly doped layer which ideally has as low a contact resistance as possible. This will also be focused on here.

The most common, and generally the only, p-type dopant used in Silicon and Germanium is boron. When doped to a sufficiently high level, typically around 1%, a SiB layer becomes metallic [96]. Heavily doped p-type contacts are useful as low resistance contacts for SPAD devices [97]. It has also been found that at extremely high concentrations (above a few percent) of boron it is possible for the layer to become superconducting [98]. Other useful applications include SPAD devices, solar cell contacts, and also in etching processes that are investigated later in this work. With regards to etching, a higher dopant concentration produces a more resistant layer to solutions such as KOH and TMAH [68]. For this particular application not all dopant atoms need to be electrically active, however as this is of

importance for other application and for the layer quality (which may be needed for subsequent growth) this aspect will still play an important role in this chapter.

An important element to realise these applications is the mass production of such devices. The ultimate industrial tool, RP-CVD, has the advantage over other tools such as MBE, because of its capability for fast and reproducible turn around. This is crucial for device production moving forward, so it is important that it is optimized with an industrial growth system.

As doping to high levels has proved hard to achieve, this has acquired notable interest. A selection of layers are presented in Table 5. Boron concentrations in excess of $2x10^{19}cm^{-3}$ have been achieved using different tools including MBE and atmospheric plasma CVD. While MBE has demonstrated some of the highest concentrations, it has the drawback of being a slow technique which may not be suitable for mass production. AP-PCVD has also shown large amounts of boron in Silicon at low temperatures, where concentrations of $8x10^{19}cm^{-3}$ have been achieved. However at this concentration the activation ratio is only around 10%, which indicates many of the boron atoms have taken up interstitial sites. Below this dopant level the activation ratio was seen to be close to 100% [31] [99]. Thermal CVD systems have also been used to try and obtain high doping concentrations, but at relatively low growth temperatures (around $650^{o}C$) the activation ratio is seen to reduce significantly (without further UV treatment) once the boron concentration exceeds the relatively low value of $1x10^{18}cm^{-3}$ [100]. While these methods have mostly been shown to successfully produce high boron concentrations in Silicon, very little work exists which reports high levels of boron obtained specifically using RP-CVD.

While much of the growth using CVD systems may be carried out at temperatures approaching $1000^{o}C$, for the production of SiB layers this is not appropriate. Both large amounts of diffusion and poor crystal quality obtained while doping heavily at high temperatures mean that any results are likely to be completely impractical from a device perspective. Although it is possible to force more dopant atoms into a layer at higher temperatures, a larger number in a statistical column does not necessarily equate to an improvement in device quality. This has also been noticed

using an MBE system, which achieved a doping concentration of $6x10^{20}cm^{-3}$ but had poor crystal quality [102].

Table 5: SiB concentrations, from literature, obtained using different growth techniques, with observations on layer quality.

Method	Concentration (cm^{-3})	Crystal Quality	Ref
AP-PCVD	$8x10^{19}$	Good	[99]
MBE	$6x10^{20}$	Very Poor	[102]
GS-MBE	$1.3x10^{21}$	Average, lots in interstitial sites	[31]
T-CVD	$1.5x10^{20}$	Required UV radiation to activate	[100]
T-CVD	$1x10^{18}$	Without UV radiation	[100]

A brief overview of some of the B-concentrations obtained in epitaxially grown Si layers is presented in Table 5. It can be seen that CVD struggles to match the concentrations measured in samples grown by MBE, and that a common problem is the quality of the layers, due to the non-substitutional nature of the boron incorporation. Work using UHV-CVD has shown a composition of around $1x10^{21}cm^{-3}$ measured using SIMS [101]. This layer was also noted to be polycrystalline. With this in mind it is advantageous for device development that doped layers are also of high quality, which often isn't the case [102]. To fully optimise a doped layer, all the atoms should be electrically active and the growth quality of other layers around it in the structure should not be compromised. To reduce the smearing of the dopant profiles and maintain a good epitaxial structure, lower temperatures have been used to grow samples with varying conditions controlling the boron precursor gases.

5.2 Growth of highly doped SiB layers

The structure which was grown is described below in Figure 5.1. A nominally undoped (001) substrate was used as the growth platform, and a doped Silicon layer was grown on top.

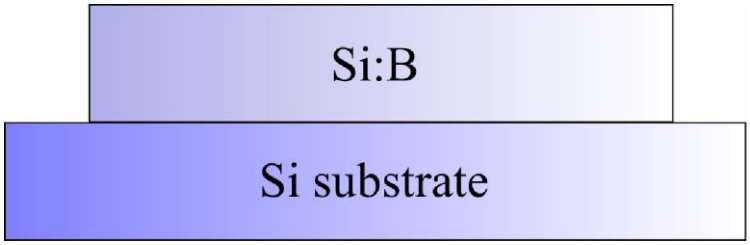


Figure 5.1: Growth plan for SiB epilayer. Doped Silicon layer is grown onto nominally undoped substrate.

A selection of SiB epilayers was grown with varying growth parameters. The samples are tabulated in Table 6.

Table 6: Table of each sample investigated with their respective growth temperature and ADC value.

Sample	Temp (°C)	ADC
12-018	700	10.2
12-019	700	30.3
12-020	700	50.8
12-021	600	2.0
12-023	800	30.3
12-270	700	50.8

The two main parameters which were altered were the growth temperature and the ADC value. This ADC value can be expressed using the equation defined by

$$ADC = 10^{\frac{\log(conc)-c}{m}}$$ Equation 5.1

Where c and m are the intercept and gradient of the plot (ADC vs conc.). The first 3 samples were all grown at 700°C, with varying ADC values for each. One sample was grown at an elevated temperature of 800°C, and another at the lower temperature of 600°C. Each of these five samples was grown to an approximate thickness of 100 nm. The final sample was grown at 700°C also, but at thicknesses much greater than 100 nm. All substrates used were cleaned using an HF dip and cleaning process to ensure ideal conditions pre growth.

This tested the limits of which temperatures would produce structures void of clustering and polycrystalline growth features. Presented in this chapter are those which contained the highest compositions. As discussed by Glass et al [31] the solubility limit (under equilibrium conditions) for boron doping reduces with decreasing temperature, so therefore unsurprisingly samples grown below 700^{o}C did not exhibit high levels of boron composition. Samples which were grown at high temperatures, which theoretically have higher solubility limits, were seen to begin undergoing a transition from perfectly crystalline to a more disrupted layer as the amount of boron incorporated into the layer increased. At the higher temperatures the Boron is likely to segregate out more, which can result in clustering and degradation of the layer as boron crystals form together instead of residing in substitutional sites. The advantage of growing low temperature Silicon is that the adatoms have less energy, so are less likely to form steps. This allows for an atomically flat surface. The temperature must also be sufficiently high so that boron atoms can be absorbed onto the surface. While work has been performed on low temperature Silicon boron layers (less than 500^{o}C) there is impracticality to this (aside from the solubility limit) [103]. At such low temperatures growth rates will limit the speed at which such layers could be produced for industrial application.

As the temperature increases the crystallinity of the layer appears to become less "perfect". Therefore lower growth temperatures are ideal so long as enough boron can be incorporated into the layer. While the solubility limit suggests that an electrically active doping concentration cannot exceed 2x10^{19}cm^{-3} there are ways to overcome this. For this work disilane was used as the Silicon precursor. At 700^{o}C the deposition rate is faster than that of other precursors. This fast growth rate means that the boron atoms may be trapped in the lower layers during epitaxy before they are able to migrate (non equilibrium growth). For equilibrium growth the layers are formed through the FvdM or VM method, as the adatoms have time to move to a surface step before subsequent layers are deposited. The premise for non-equilibrium growth is that the adatoms do not have enough time to move to a surface site before the next layer is deposited. This makes FvdM growth difficult and the onset of 3D islands more prominent. Eventually mounds of atoms will be present across the substrate and the crystallinity will be disrupted. The disilane precursor used allows for faster deposition of atomic layers. Use of this precursor can be difficult, and to achieve

perfect layers requires precise calibration. At high temperatures the solubility limit is increased, however, particulates can form in the gas phase due to the increased reactivity of the precursor, and lead to polycrystalline growth.

At the lower growth temperatures the energy of the boron atoms is much lower. It becomes more energetically favourable for the dopant atoms to form more surfaces, similar to that of 3D island growth. Also the adatoms mean free migration length on the surface may not be enough to reach a step. This can then lead to clusters of atoms growing on the surface away from step edges which can result in the polycrystalline features. When an atom does not reach a step and is deposited on the surface a number of potential wells are set up which can form sites for other atoms (without enough energy) to fall into. This is another reason for using suitably high growth temperatures where the migration length is greater.

5.3 Structural characterization of SiB Layers

Initial characterization of the SiB layers was structural, including TEM, AFM, and XRD. Unfortunately SIMS analysis was not utilised for this analysis. This is because at high concentrations such as the ones expected in this chapter the data is unreliable because there are few/no references to obtain the doping profiles. The limit quoted for EAG SIMS analysis is up to 1%.

Several samples were analysed which were grown via the different methods described in the previous section. TEM images of a selection of the samples can be seen in Figures 5.2, 5.3, and 5.4.

As can be seen for samples 12-019 and 12-020, the crystal quality is perfectly fine, and the thicknesses of the layers are both approximately 100nm. Sample 12-018 is analogous in both crystal quality and thickness also.

It is clear from the TEM image of sample 12-0270 that the epilayer quality is non optimal. There are a few possible reasons for this: The first is that the layer is beginning to become polycrystalline with such a high level of boron composition, the second is that there is a clustering of boron atoms which is leading to large levels of localised strain.

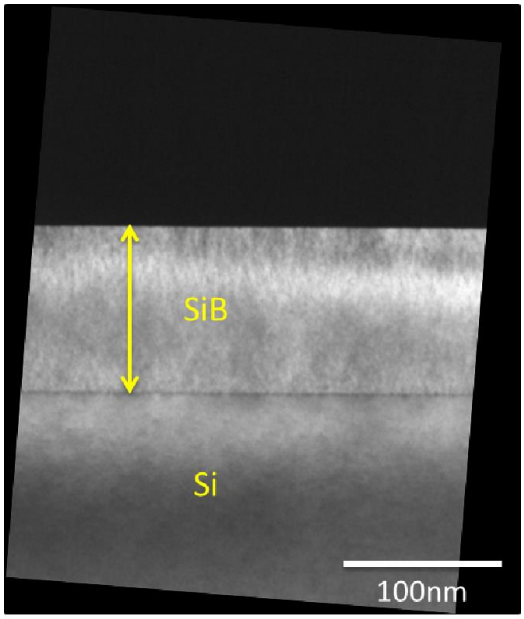


Figure 5.2: Dark field (220) TEM image of sample 12-019.

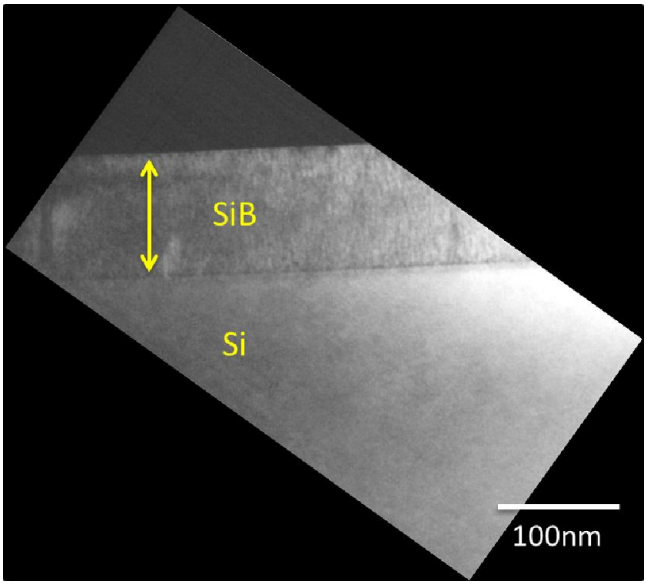


Figure 5.3: Dark field (004) TEM image of sample 12-020.

Alternatively, due to the high level of boron present and the thickness of the layer, the SiB may be starting to relax towards a new lattice parameter. This assumes that the layer is still crystalline. And therefore has not become poly at this point. With such a high doping level the layer is to be considered an alloy. Therefore it could be assumed that the layer has been grown to at least its critical thickness. If the layer was grown beyond this limit it would be reasonable to expect dislocation to form during the relaxation process. As these are not visible in the TEM image, the

critical thickness has not been exceeded. At the lower doping concentrations the critical thickness would be higher, and therefore thicker layers could be grown if required. With TEM alone it is difficult to fully understand the cause of the problem, therefore further analysis is required to try and make a conclusion.

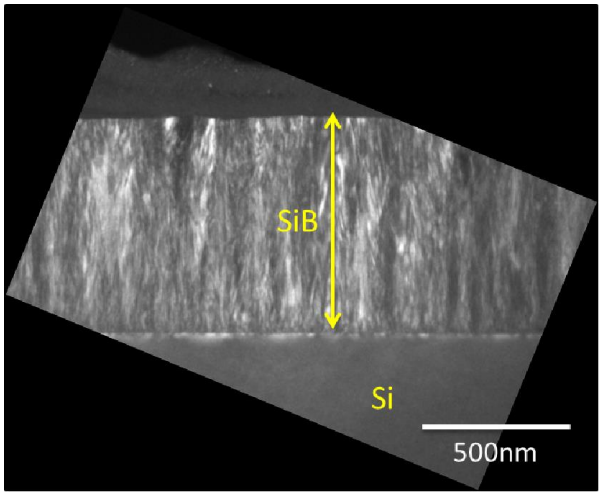


Figure 5.4: Dark field (220) TEM image of sample 12-270. On set of polycrystallinity/increased strain in the epilayer can be observed.

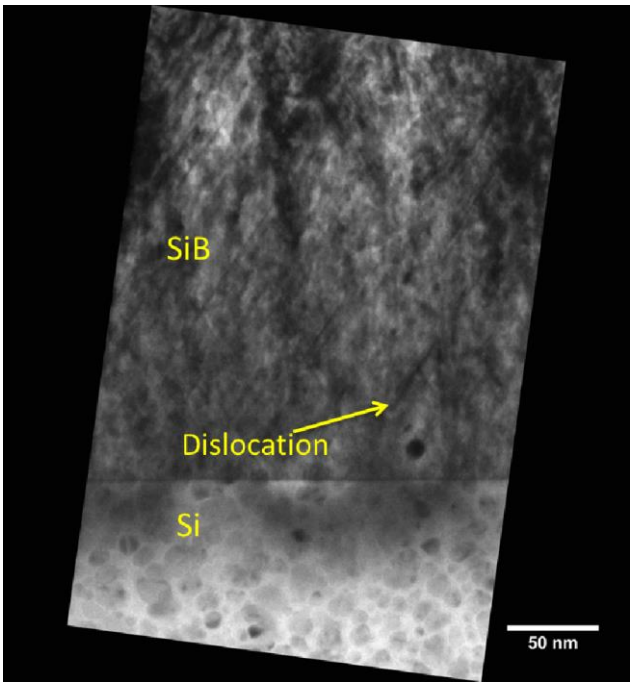


Figure 5.5: Dark field (220) TEM image of sample 12-135. Due to a thinner critical thickness dislocations can be seen to be forming in the layer.

To test this hypothesis another sample was analysed. The growth of which was similar to sample 12-270. For this sample (12-135) the thickness and growth temperature was identical to that of sample 12-270, but the dopant gas flow was increased. This would have two potential effects on the layer. The first of which would be an increased chance of interstitial doping, while the second would be a reduced critical thickness. As the lattice parameter of the layer should be smaller than that of sample 12-270, due to a higher boron concentration, the increased difference between that and the Silicon substrate will cause a reduction in critical thickness. This has resulted in the formation of dislocations in the layer, observed in Figure 5.5. Therefore it would be reasonable to conclude that in this case relaxation has begun, as opposed to interstitial clustering. It is however possible that interstitial clustering is also present. This could be the case just below the dislocation where a dark circle can be observed. This sample highlights the precise control necessary to obtain crystalline layers, and also the sensitive nature of all the growth parameters and conditions.

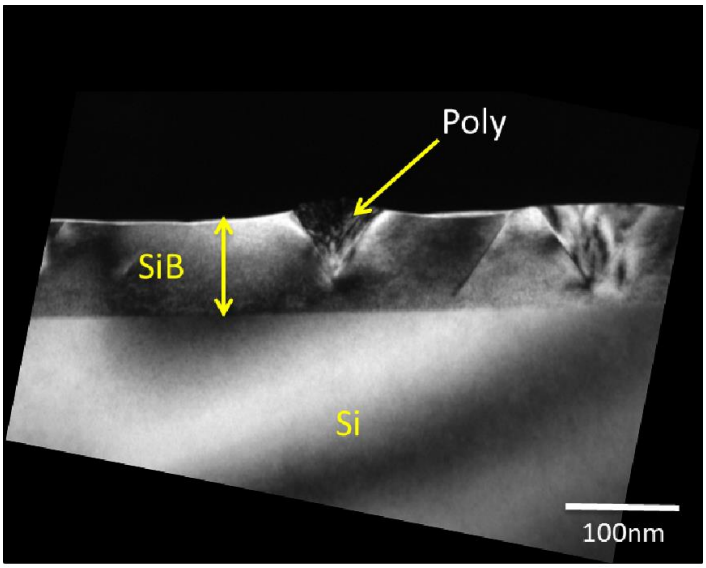


Figure 5.6: Dark field (220) TEM image of sample 12-016. Polycrystalline mounds can be seen around the dark areas in the epilayer.

Interestingly sample 12-016, which is not investigated fully in this work, displayed some unique features, shown in Figure 5.6. This sample was grown at the lower temperature of 600^{o}C with an ADC of approximately 10. At these conditions polycrystalline growth has begun, which is seen as the inverted pyramids on the surface, accompanied by stacking faults. It can be noticed that there are portions of

the feature which cannot be seen fully, which is due to its polycrystalline nature. This is likely described by the clustering of Boron atoms at the low growth temperature and a resultant 3D island growth. This can be described by adatom transport. The temperature applied to the substrate gives the adatoms energy to traverse the surface. The migration length, which is of the order nm, of the species on the surface is increased with temperature. The bonding of boron-boron or Silicon-Silicon is more energetically favourable than boron-Silicon bonds. This means at lower temperatures where the adatoms have less energy, boron is likely to cluster together, which could potentially lead to the loss of crystallinity [104].

The surface morphology of the structures was investigated through AFM to highlight the quality of the surfaces. Samples 12-018, 12-019, and 12-020 (Figure 5.7b) all displayed very smooth surfaces, with a sub nanometre RMS roughness. As the mismatch between the layers is incredibly small a smooth surface is unsurprising. When different growth temperatures were used the surface quality was seen to change. Sample 12-023, which was grown at 800^{o}C, displayed a much rougher surface as seen in Figure 5.7.

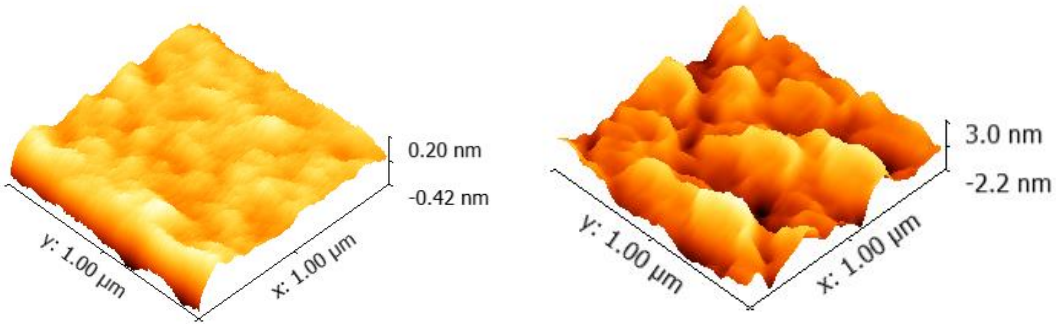


Figure 5.7: 3D AFM scan of the surface of samples 12-020 (a) and 12-023 (b).

It is speculated that this roughening could be due to a clustering effect. Due to the large level of boron in the layer it is possible that the dopant atoms have been able to segregate out at the higher growth temperature and begun to form small clusters of boron, which has disrupted the layer quality. TEM of the sample, shown in Figure 5.8 highlights the roughening of the surface.

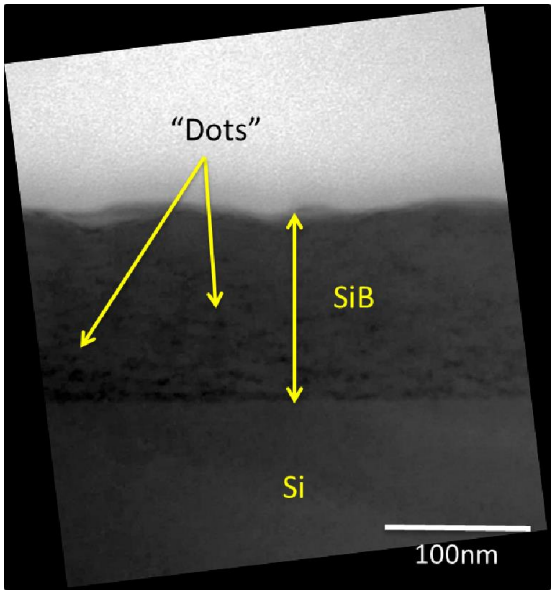


Figure 5.8: Bright field (220) TEM image of sample 12-023. Small dark dots in the epilayer are thought to be clusters of atoms or areas of polycrystallinity.

Around the interface a series of dark dots can be seen, which could potentially be a sign of boron clustering. This early disruption in the growth could have filtered up to the surface, therefore resulting in the wavy surface. The increased temperature could lead to the increased migration of the boron atoms. An alternative reason for the wavy surface could be attributed to particulates forming in the gas phase and their deposition forming a non flat platform for further growth.

Table 7: The RMS surface roughnesses of each SiB sample.

Sample	RMS roughness (nm)	Temp (^{o}C)
12-018	0.05 +/-0.01nm	700
12-019	0.05 +/-0.01nm	700
12-020	0.06 +/-0.01nm	700
12-021	0.06 +/-0.01nm	600
12-023	0.90 +/-0.1nm	800
12-270	1.23 +/-0.1nm	700

The RMS roughness values for each sample are presented in Table 7. An increase in roughness can be seen in the two thick samples grown at 700^{o}C. The TEM image of sample 12-270 shows a mottled effect throughout the epilayer, which it is

thought could be the reason for the roughness increase in this sample. The clustering or localised strain experienced through the whole layer may have led to the RMS roughness rising to over 1nm.

5.3.1 Dopant concentration calculations from XRD

Simple rocking curves were performed on the samples to calculate the doping density and also verify the thicknesses of the layers by calculating the periodicity of the fringes present on the Silicon peak of each scan. Doping densities could be calculated by measuring the angle for the SiB peak on the rocking curve to obtain the lattice parameter $d = \frac{a}{\sqrt{h^2+k^2+l^2}}$. For this data (004) rocking curves were taken, meaning that the lattice parameter could be found using:

$$a = \frac{2\lambda}{sin\theta} \qquad \text{Equation 5.2}$$

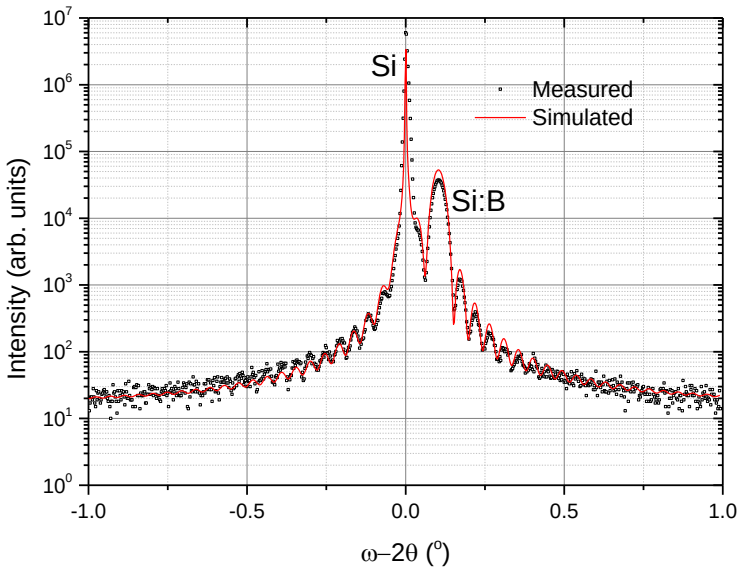


Figure 5.9: (004) Rocking curve for sample 12-020 with fitting curve (red). SiB peak can be observed to the right of the Silicon peak accompanied by thickness fringes.

The lattice parameter for the SiB layer is related to the Boron concentration N_B by

$$\frac{a_{SiB}-a_{Si}}{a_{Si}} = \beta N_B \qquad \text{Equation 5.3}$$

[105], where $\beta = -5.6 \times 10^{-24} cm^3$, from which the boron concentration can be obtained.

The rocking curve in Figure 5.9 is an example of one of the curves used to calculate the thickness and doping density. The simulated curve clearly shows the extra SiB peak and thickness fringes used to calculate the boron concentration and epilayer thickness, respectively.

Table 8: Thickness of the SiB layer in each sample evaluated from TEM and XRD measurements.

Sample	Thickness TEM (nm)	Thickness XRD (nm)
12-018	121 +/- 3	120 +/- 5
12-019	118 +/- 3	118 +/- 3
12-020	116 +/- 3	118 +/- 4
12-021		
12-023	130 +/- 3	130 +/- 4
12-268	498 +/- 3	498 +/- 6
12-270	720 +/- 3	723 +/- 5

A comparison of the layer thickness measured for each sample is presented in Table 8. Thicknesses measured by TEM and XRD are in very good agreement only differing by a few nanometres at most. Sample 12-268 was a sample which was not heavily Boron doped and therefore did not present a SiB peak, shown in Figure 5.10. Only at sufficiently high doping levels, approaching alloys, will a Bragg peak be noticed with standard lab based x-ray kit. For sample 12-270 there is a clear Bragg peak, however it is not as defined as some of the other samples. A broadened peak is often the sign of imperfect growth, which could be caused by dislocations or the onset of poly growth. The relaxation process between two layers creates a mosaicity in the layer due to the growth occurring at slightly different planes. The result is that diffraction at the Bragg angle is slightly altered which causes an broadening of the peaks. In this case, when corroborated with the TEM, it can be suggested that a clustering of boron atoms or relaxation processes could be the reason.

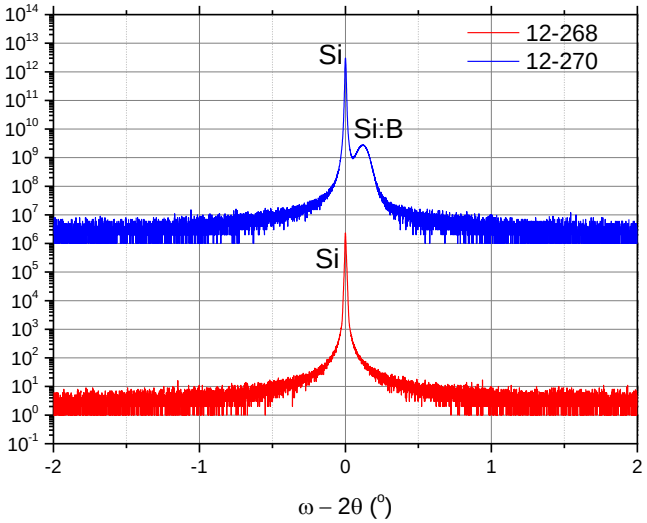


Figure 5.10: (004) rocking curves for samples 12-268 and 12-270. Sample 12-268 was a sample with essentially intrinsic doping for comparison.

Table 9: Boron compositions calculated from (004) XRD rocking curves.

Sample	Doping Conc. XRD (cm^{-3})
12-018	$2.60x10^{20}$
12-019	$3.99x10^{20}$
12-020	$4.50x10^{20}$
12-021	
12-023	$2.40x10^{20}$
12-268	$<10^{17}$
12-270	$5.36x10^{20}$

Analysis of the Bragg peaks from the rocking curves enabled the calculation of the Boron concentration for each sample, which is tabulated in Table 9. The noticeable anomaly is that of sample 12-268 which did not present an additional Bragg peak, due to its intrinsic nature. The remaining samples all showed very high levels of boron doping, approaching 1%, this is matrix level as opposed to dopant level, which has not been reported before using RP-CVD. This result is significant as RP-CVD is considered the ultimate industry tool, and has therefore been shown to be capable of producing near alloy level crystalline SiB layers on a mass scale.

Importantly they can be incorporated onto monolithic or thick multi-layered structures, which are generally not possible to obtain in large quantities through MBE. The uncertainty on these calculations comes mainly from the value of β used to find the final doping concentration. This is an experimental value and carries a fair amount of uncertainty. Based on some of the values reported in literature there is at least a 28% error on this value [106][107].

Prior structural characterization has confirmed the excellent quality of the epilayers, which is often a concern when trying to incorporate large quantities of foreign atoms into substitutional sites. The thicker sample, 12-270, showed some signs of clustering at a concentration just above 1%, however the thinner samples displayed no such issues. This is advantageous over other doping techniques such as ion implantation where the layer is often subject to damage, and annealing steps are required to incorporate the atoms.

5.4 Electrical characterization of SiB layers

The SiB samples were also characterized electrically to support the structural characteristics. The main analysis was performed through VdP measurement to find the carrier concentration, which could then be compared to those values found using the rocking curve method. For each method the electrically active doping concentration is found. For the XRD method only dopant atoms in lattice (substitutional) sites will contribute to the diffracted signal, and similarly for the hall effect measurements only substitutional atoms will contribute to the hall voltage. Any atoms in interstitial sites will not contribute in either method. Although technically this will provide a minimum density of boron atoms in the Silicon, any atoms not in a diamond lattice position will be electrically inactive and therefore provide no benefit from a device perspective. As the layers have significant doping levels, greater than the solubility limit, interstitial doping is inevitable, which can lead to a reduction in layer quality through clustering.

The sheet resistance for each of the seven samples under investigation was measured with increasing temperature. This was used to identify the nature of each sample. The plot, in Figure 5.11, shows that all the samples, except for 12-268, show a metallic behaviour in their sheet resistance (i.e. a low resistance that increases slowly towards higher temperature). This is expected as the doping

density is sufficiently high. Sample 12-268 shows an intrinsic semiconductor trend with its sheet resistance decreasing from a high value as temperature increases. This supports the XRD data which showed no sign of a separate SiB peak. This confirms that this sample is essentially intrinsic, and has been used as an example which displays semiconductor behaviour.

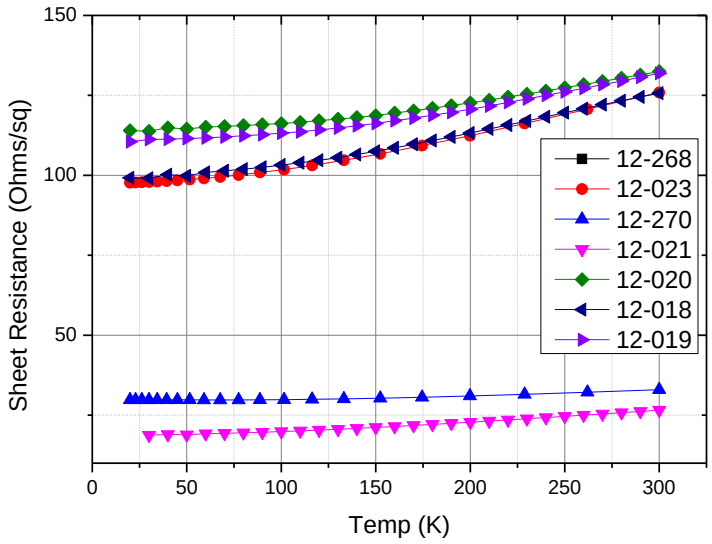


Figure 5.11: Plot of sheet resistance for each SiB sample. Measurements are taken at regular intervals of temperature .All samples are highly boron doped, except for sample 12-268, which was essentially intrinsic, therefore showing semiconductor resistance nature. Sample 12-268 is not shown on this plot due to its much higher resistance compared to all other samples.

5.4.1 Dopant concentration calculations from VdP

To find the carrier concentration throughout each SiB sample the sheet density was measured using the VdP set up. Once this data was obtained the carrier concentration was calculated by dividing by the thickness of each sample. Figure 5.12 shows the boron composition for all the samples.

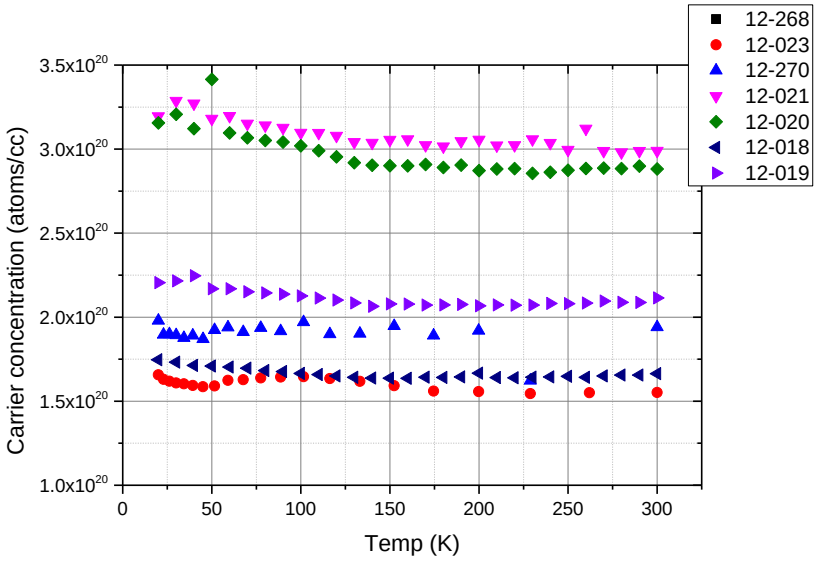


Figure 5.12: Carrier concentration for each SiB sample, measured by VdP. Data is excluding sample 12-268, which had a carrier concentration much lower than other samples.

The two samples which displayed the largest boron concentration were samples 12-019 and 12-020, with over $3x10^{20}cm^{-3}$. This value is well above the proposed solubility limit of boron at $2x10^{19}cm^{-3}$ at 700°C [31]. Despite this, as the TEM shows along with the AFM, the samples are monocrystalline and have no signs of any serious growth issues. The other samples in this batch also show high levels of boron concentration, with all above $1.5x10^{20}cm^{-3}$. The calculated values for doping density are tabulated in Table 10.

Table 10: Doping concentrations for each sample measured through Van der Pauw technique.

Sample	Doping Conc. VdP (cm^{-3})	Temp (°C)
12-018	$1.66x10^{20}$	700
12-019	$2.11x10^{20}$	700
12-020	$3.00x10^{20}$	700
12-021	$3.08x10^{20}$	600
12-023	$1.60x10^{20}$	800
12-270	$1.89x10^{20}$	700

5.5 SiB bulk structural and electrical summary

This chapter has investigated high concentrations of boron in Silicon layers, while maintaining crystal structure. Above the solubility limit dopant atoms begin to occupy non-substitutional sites and clustering can become a problem. As boron does not form a diamond structure (rhombohedral) when clusters of another structure are formed the crystal quality is disrupted. AFM and TEM were useful in inspecting any potential growth issues, which for most samples were not observed. Both XRD (004) rocking curves and Hall Effect measurements confirmed the electrically active boron concentration in the layers. Sample 12-020 was shown to have a concentration approaching 1% which is the highest concentration reported by RP-CVD. Importantly these measurements have reported the concentration of boron atoms in substitutional sites, as opposed to the total in the layer. Large amounts of interstitial atoms will play no role in the device, while potentially disrupting the layer quality. Although energy dispersive spectroscopy (EDS) could have been used, it struggles to detect the light elements due to the low energy peaks occurring close to the electronic noise, and a low yield of x-rays.

Table 11: Summary table of the key parameters used and measured, for SiB layers.

Sample	Temp (celcius)	ADC	RMS roughness (nm)	Thickness TEM (nm)	Doping Conc. XRD (cm^{-3})
12-018	700	10.2	0.05	121 +/- 3	$2.60x10^{20}$
12-019	700	30.3	0.05	118 +/- 3	$3.99x10^{20}$
12-020	700	50.8	0.06	116 +/- 3	$4.5x10^{20}$
12-021	600	2.0	0.06		
12-023	800	30.3	0.9	130 +/- 3	$2.40x10^{20}$
12-268	700	200		498 +/- 3	Too Low
12-270	700	50.8	1.23	720 +/- 3	$5.36x10^{20}$

Comparing the concentrations obtained in this study to those described in Table 5 does not show the highest Boron concentration reported. However the results do present superior growth quality for concentrations which have not been reported for

RP-CVD as the growth technique. The quality of growth is also improved when compared to most other methods, particularly of comparable dopant concentrations.

To summarise the optimal conditions for highly doped SiB layers in an RP-CVD system Table 11 highlights the key findings. Highly electrically active and accurately doped layers are important for top and bottom contacts in SPAD devices, and may be required to be produced on a large scale. The proof that this doping level can be achieved using an RP-CVD system demonstrates the capability for industrially developing such layers or structures, which cannot be done so easily with research specific tools.

A key result for these highly doped layers is their potential for suspending a structure, which is explored in the next chapter. A highly doped layer can offer an alternative as an etch resistant layer to using a different material such as Germanium. Paramount to these samples was that the doping level was high enough to allow for wet etch resistance (which is properly explored in Chapter 6), while maintaining an excellent crystal quality, which can be compromised by large amounts of interstitial doping and clustering of dopants. The smoothest samples were produced when using a temperature of 700^{o}C. The lower temperature sample could suffer from some boron clustering which would not be ideal, especially when growing a thicker layer. An increase in the ADC value shows an increase in doping concentration up to the value of 50.8. A thicker layer of this sample (12-270) shows a build-up of strain or the beginnings of polycrystalline growth, which shows a thickness limit, should be imposed. This layer is still crystalline as seen in the diffractive imaging, however it is thought that much thicker than this may result in a defected layer. For application to suspending structures, if the layer is completely resistant to an etchant, a thick layer would not be necessary. At high boron concentrations, especially beyond the solubility limit boron-boron, clustering is more prominent. As there are more boron atoms there is a higher probability of them bonding together, purely through probability and the lower energy required to bond the atoms.

While the value of β is not well defined, value obtained for bulk lattice parameter and doping concentration have been used to show a value for this work, presented in Figure 5.13 (where doping concentration obtained from bulk measurements are

plotted with the lattice parameter change for each sample). By using the doping concentration obtained through XRD and hall measurements the value was found to be -5.52+/-0.55x10^{-24}cm^3 and 6.97 +/-2.31x10^{-24}cm^3 respectively. These values are consistent with those published, and may be catalogued amongst them.

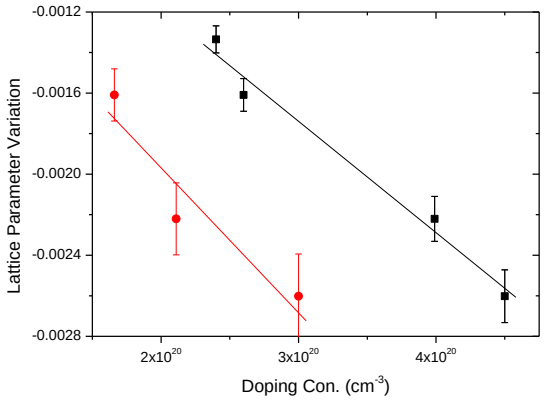


Figure 5.13: Plot describing β values obtained for Equation 5.1. Red data represents hall data, and black shows XRD. Gradient of each fit gives β value.

Table 12: Summary of boron doping in Silicon, including data obtained from this work. Crystal quality is also described due to its importance for device application.

Method	Concentration	Crystal Quality	Ref
AP-PCVD	8.00E+19	Fine	[99]
MBE	6.00E+20	Very Poor	[102]
MBE	1.30E+21	Ok, lots in interstitial sites	[31]
T-CVD	1.50E+20	Required UV radiation to activate	[100]
T-CVD	1.00E+18	Without UV radiation	[100]
RP-CVD	**4.50E+20**	**Good crystallinity**	**This work**

The highest doping concentration reported is compared to other techniques in Table 12. It is evident that this work's result is the highest concentration of boron obtained through the CVD technique. Most importantly the quality of the layer is

superior to other layers doped using different techniques. This layer is therefore ideal for a platform for further growth. A subsequent structure could be grown on top of the Silicon boron layer before further fabrication. This is particularly useful for potentially suspending Silicon without having to grow it on a mismatched substrate or oxide.

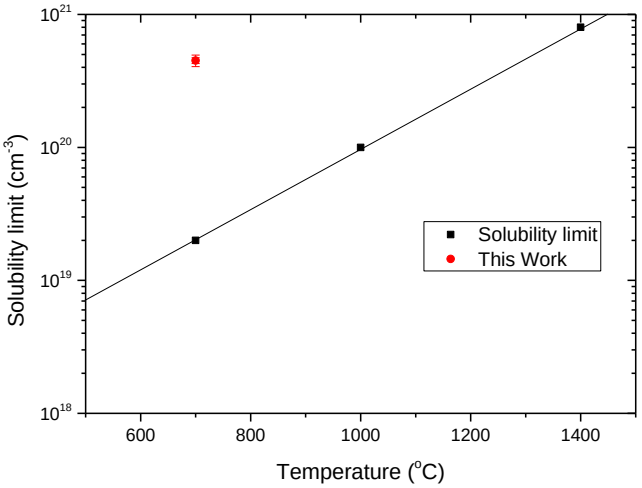


Figure 5.14: Plot demonstrating the results from this work compared to solubility limits at different growth temperatures. The red data point shows an order of magnitude increase to the doping concentration above the solubility limit at 700°C.

While the doping concentration could potentially be increased with a higher temperature, which corresponds to a higher solubility limit, the structure quality could be compromised. The elevated temperature would lead to a more significant migration of the dopant atoms. From a device perspective, a smeared doping profile with a higher number of dopants occupying regions intended to be intrinsic, is far from ideal. This result is also compared to the reported solubility limits at different growth temperatures in Figure 5.14. It can be seen that the concentration of boron in the layer grown in this work exceeds the solubility limit at 700°C by more than an order of magnitude. This is presumed to be caused by the increased growth rate of Silicon when using disilane as the precursor. The boron atoms may be “trapped” by succeeding layers before they are able to move too far or form clusters.

Demonstrated here is a high doping concentration which can be localised to a specific region in a structure. Also from the point of view of an etch barrier, a structure could be grown on top (or incorporated into) without jeopardising the other doping profiles in the structure. The growth of such layers, for either application, can be performed quickly in bulk with relatively small expense in one system. This allows for new devices making use of the qualities of such layers possible on an industrial level. The superior growth quality of such layers provides an appealing incentive for growing high doped Silicon boron layers in the RP-CVD system.

Chapter 6

Suspended Membrane and Wires

6.1 Development of membranes and wires

In this experimental chapter several different suspended structures are investigated. SiB membranes are fabricated using a backside etch to remove a square of substrate, leaving a SiB layer of approximately 100nm. SiB wires were also designed and fabricated using a top side etch approach. Multiple thin strips of SiB were suspended away from the substrate analogous to bridges. Micro-diffraction was employed at the DIAMOND light source to characterize both the membranes and the wires. Because a boron diamond crystal (which is fictitious) would have a smaller lattice parameter than Silicon, a doped layer grown on a Silicon substrate would be tensile strained. The lattice will be slightly stretched initially while connected to the substrate. Upon suspension it is interesting to observe what happens to the layer when the substrate no longer holds the layer in place.

Samples used for this chapter are the same as those characterized in the previous investigation, and were selected accordingly based on their suitability to resist the etchant used, and quantity remaining.

6.1.1 Motivation for suspended structures

While SPAD devices may be grown on SOI substrates in the future, this work has concentrated on structures grown on Silicon substrates. By removal of this parasitic substrate with an etching process, unwanted leakages and unnecessary resistance can be eliminated. This could lead to a suspended SPAD device which incorporates an epitaxial bottom contact layer.

The fabrication of any suspended structures required an etching process to remove sections of substrate or epilayer. As described in Chapter 2 at high enough levels of boron doping the layer starts to transition from etch resistant to essentially an etch stop. At the lower concentrations the etch rate is rapid, and those samples are not suitable for etching. The etch rate decreases steeply at boron concentrations around $10^{20}cm^{-3}$. The samples containing the highest dopant concentrations, which were measured using XRD and hall effect, should be suitable as an etch stop. Flat suspended structures are ideal for any further growth which may be necessary for a device, so minimal tilt is important. It is also important that the etch resistance is

suitably strong so that any etching post growth will not damage epilayers in a device structure.

A selection of samples, including those with the highest B-composition, were tested in a TMAH solution to discover how etch resistant the doped layers were. The TMAH solution was set at a temperature of 80^{o}C, and the samples were immersed in the etchant for range of different times. The samples were placed into the TMAH bath without any surface protection, so were therefore subject to etching immediately. Although this was not identical to having a membrane this was useful to simulate how a membrane would survive without having to spend upwards of 12 hours etching the substrate for each sample. Also it was useful to explore the durability of the layers for wire fabrication, as this required a 90 minute etch from the surface instead of from the back through the substrate. After each period in the bath the samples were analysed using thickness fringes from XRD to obtain the thickness of the SiB layer.

Four different samples were measured in this study shown in Figure 6.1. Samples 12-020 and 12-019 both had doping levels around the level which is expected to act as an etch stop. Both other samples (12-018 and 12-133) had boron compositions lower than that, and were therefore expected to be etched away. This can be observed for both samples which are etched away after less than 500 seconds in the bath.

Interestingly, sample 12-019 also started to be reduced in thickness after just over 500 seconds in the bath, despite having a high doping concentration. A rocking curve of several pieces of wafer from 12-019 were analysed. This showed that there was a slight variation in doping concentration over the wafer itself with slightly higher doping concentrations at the edge. Measurements for the electrical characterization used pieces of wafer closer to the edge due to the fact that many other central pieces were used for other analysis and characterization, including the etch study. It is suggested that this central piece had a slightly lower doping concentration than that required to be an etch stop which is why it did not survive long in the TMAH.

Sample 12-020 shows a much more definitive resistance to the etchant over a long period of time, where the thickness remains essentially the same within the

experimental uncertainty. Therefore it can be concluded that this sample is the most likely to survive the membrane and wire fabrication process, and could also be used as an etch stop in the future. This would have the advantage over introducing a Germanium layer if growing a Silicon epilayer, as there would be a minimal lattice mismatch that would create misfit and threading dislocations if using Germanium.

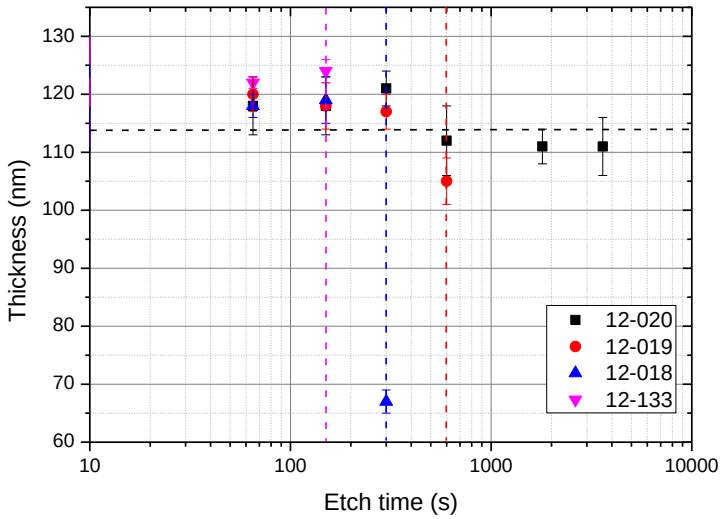


Figure 6.1: Thickness calculations from (004) rocking curves for each sample against etch time. When the sample is etched away (or becomes too thin) no thickness fringes are observed. Therefore the sample is considered not etch resistant (shown by the dashed line).

Sample 12-018 shows the thinning of the SiB epilayer. The layer can be seen to have reduced to 67 nm after 300 seconds of etching before disappearing all together after the 10 minute etch time. Further intermediate etch times would be required to estimate the etch rate. The main conclusion from this small investigation is the suitability of sample 12-020 as a candidate for studies involving long etch times. This was the primary sample used throughout the suspended work.

6.2 Optimization of fabrication processes

The following section describes the fabrication process for both the membrane and wires. Fabrication, and associated optimisation of the process, was performed in

house. Several months were spent in an attempt to make the process more efficient due to the delicate nature of the suspended structures, and their tendency to break.

6.2.1 Membrane fabrication

A small square (approx. 1.5cm^2) was cleaved. Prior to the fabrication process the sample was cleaned in acetone and HF to remove any particles and oxides from the surface. Any remaining moisture was removed via baking on a hot plate. The first stage of fabrication was performed on the topside of the wafer, so that the SiB epilayer was protected from the etch. This essentially involved coating the surface of the epilayer with Protek to prevent any etching from the topside.

The sample was then placed into the spinner and covered by a primer solution before being spun at 2500 rpm for 60 s. This was followed by a soft (110^{o}C) and hard (220^{o}C) bake. Once back in the spinner the sample was coated in PGMEA (used as a surface treatment) and spun at 2000 rpm for 20 s. This was immediately followed by applying Protek to the surface and performing two spinning programs consecutively (both 4000 rpm for 60 s) [108].

The sample was then baked at 110^{o}C for two minutes and exposed to UV radiation in an MJB3 mask aligner. After exposure the sample was subject to a soft and hard bake. This acted to ensure the polymers in the Protek layer were resilient to the developer.

For the underside of the sample a very similar approach was taken. For this stage the membrane mask was applied during the UV exposure to form a series of squares which would become membranes. An example of one of these squares is shown is Figure 6.2. Also prior to the final hard bake the sample was developed in the ethylathtate. This removed the Protek from the membrane location so that the Silicon substrate would be immediately exposed to the etchant. The final stage after the subsequent hard bake was to remove the oxide from the membrane window in an ICP etcher.

Silicon has an etch rate in 25% TMAH of 0.45 ± 0.02 μm min^{-1} when at 80^{o}C, and was used to remove the substrate [109]. Once the sample was completed it was immersed in a bath of TMAH for 18 hours, which should remove roughly 500 μm of Si, which is the thickness of the substrate. TMAH has been used as an etchant at

temperatures ranging from 60-90^{o}C, where 80^{o}C was used for this experiment (giving an appropriate etch rate). The (111) planes are the most resistant in Silicon, which is why there is a slight narrowing through the etch pit [112] shown in section 2.8.2.

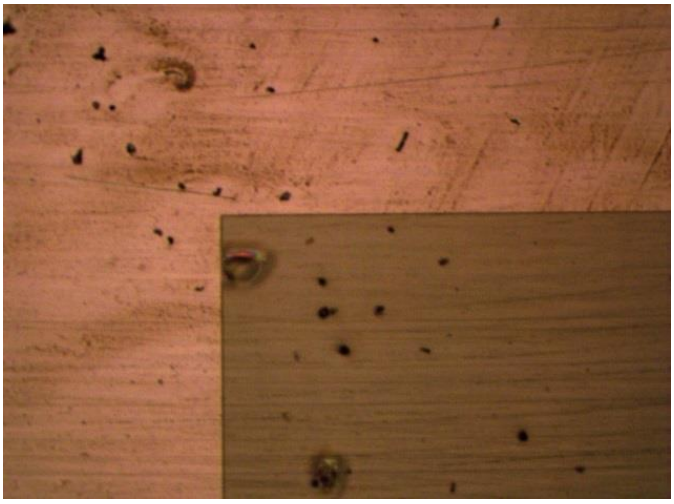


Figure 6.2: Optical image of defined membrane square after photolithography process and pre etching Membrane fabricated were 1mm x 1mm in size.

The trick to etching is to use a material as an etch stop so that the etching can be quenched. With a sufficiently high level of boron doping in the epilayer the etching should stop at this point. Based on Figure 6.1 the sample should be unharmed for a window of up to a couple of hours. Timing of the removal was therefore important, but did not have to be exact, which meant that it could be ensured that the entire substrate was removed. To check that the layer was suspended the membrane itself would produce a red tint when held against a white light source, as the red light could be transmitted (other wavelengths are absorbed).

Producing a defined square did prove to be tedious throughout this process. Initially there was an under-etching problem on the backside. Before the optimization the HF dip step was not performed, which allowed an oxide layer to form. The TMAH etch presumably etched the oxide layer between the backside and the Protek. This allowed for areas initially protected to be etched so that a square was not formed. To solve this problem the HF dip was performed to remove the oxide layer. Also by removing other particulates from the surface issues with a lack of adhesiveness were also reduced.

6.2.2 Wires fabrication

The process for developing sets of SiB wires followed a similar route; however, it was a much faster and successful method (due to the number of wires produced on each sample). Instead of the backside etch approach, the wires were formed through a top down method. The top side of the sample was protected using photoresist. A mask with the desired pattern was imposed using the same method as for the membranes. The mask used consisted of several sets of wires, with each set spanning a range of widths for each wire. The following equation describes the limit for resolution for the photolithography process:

$$R\ \alpha\ \frac{\lambda}{NA} \qquad \text{Equation 6.1}$$

Where $NA = nsin\theta$. The latest values for NA is approximately 0.9. This gives a typical resolution, based on UV radiation, of 50 nm. This is more than suitable for this investigation where the spot size of the synchrotron x-ray beam is around 1 μm, and therefore would not be able to measure wires if their size was around the resolution limit.

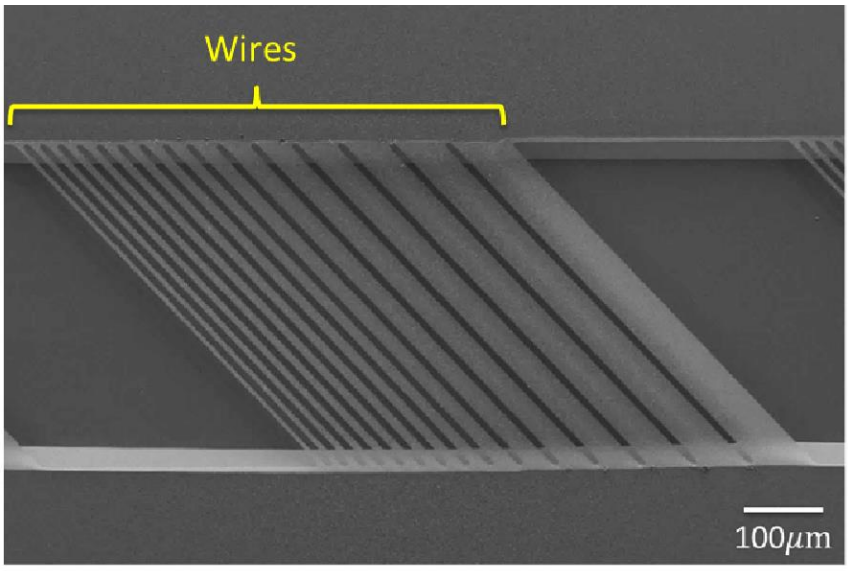


Figure 6.3: SEM image of a set of fabricated SiB wires, according to the process described in section 6.1.1.2, used for synchrotron measurements.

The samples were then bathed in a TMAH solution at 90^{o}C for 90 minutes. This was long enough to etch the unprotected areas of the surface and create the

suspended wires. A series of wires ranging from 50 μm in thickness down to only a micron or so were produced using this mask.

The result of the etching process can be observed in Figure 6.3. A clean set of wires of varying thickness was successfully produced by this method for micro XRD measurements. Areas of the wires could be mapped out and analysed for strain and tilt.

6.3 Synchrotron measurements

The suspended structures were analysed at the Diamond light source using beamline B16. The data analysis was carried out according to the method described in detail in section 3.3.2. In summary (004) RSMs were produced for each spacial position, and then maps containing information about lattice parameter and tilt could be obtained.

6.3.1 Analysis of suspended SiB membrane

The corner of the SiB membrane was mapped out to show any changes to the lattice parameter. The bulk value, away from the corner, could be assumed to be that of unsuspended SiB, and therefore could be used as a reference. The membrane can be observed in Figure 6.4, in the bottom right corner. Around the edge of the suspended structure there is an increased lattice parameter, which can be seen to decrease slightly when looking towards the centre of the membrane. However, even within the more central region, the lattice parameter is larger than that of the bulk material. As the map does not display the actual centre of the membrane it is hard to definitely conclude what is happening. If the trend of the decreasing of the lattice parameter continues the centre of the membrane could have a lower lattice parameter than the bulk. It is speculated that this could be the case, as this was observed for a similar study investigating a Germanium membrane [82].

The important information to investigate was not necessarily the absolute values of lattice parameter, but the percentage change between the bulk material and suspended structure. Generally the variation from bulk to suspended material is incredibly small; with the maximum deviation across the map only 0.013%. The edge to central difference is more than half of that compared to the bulk.

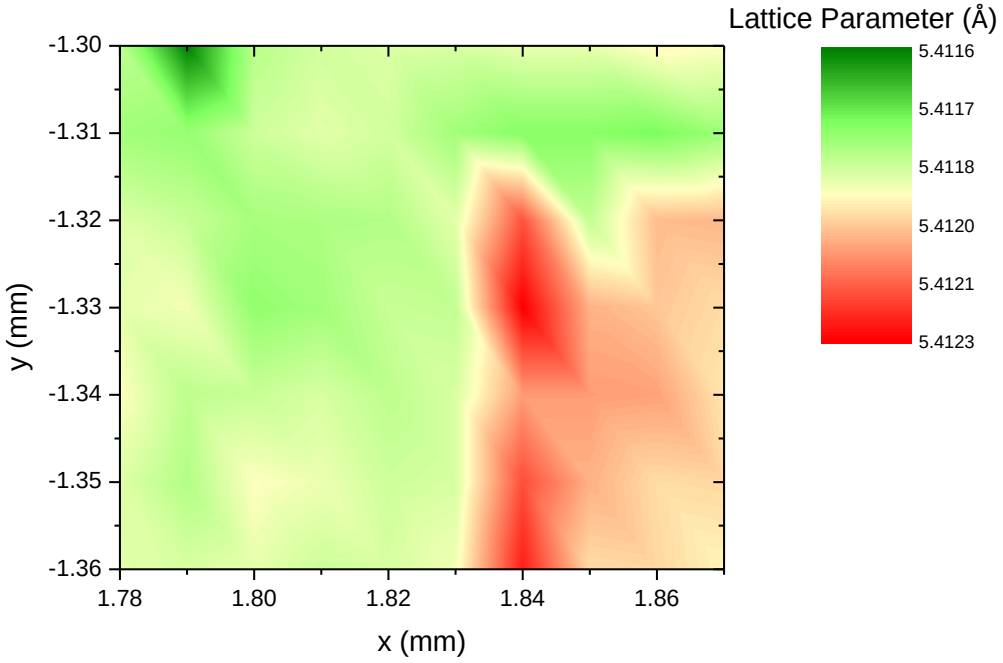


Figure 6.4: Out-of-plane lattice parameter plot from across the membrane, using (004) reflection calculations. The red area represents the corner of the suspended membrane, and the green region is the bulk SiB material. The lattice parameter is at its largest at the membrane edge before decreasing slightly towards the middle.

6.3.2 Analysis of suspended SiB wires

The SiB suspended wires were analysed using the (004) reflection to measure the out-of-plane characteristics. The micro diffraction process was performed in the manor described in section 3.3.1. A series of maps measuring the strain and tilt across the wires was produced and is presented in the following section.

The tilt across the wires can be seen to be relatively small. It can be observed that the degree of tilt is of the order 10^{-4} degrees. The maximum tilt seen across this set of wires is 8.6×10^{-4} degrees, which is considered to be negligible. This is an important result as this shows that the wires are essentially flat, and therefore can be suitable for future device fabrication.

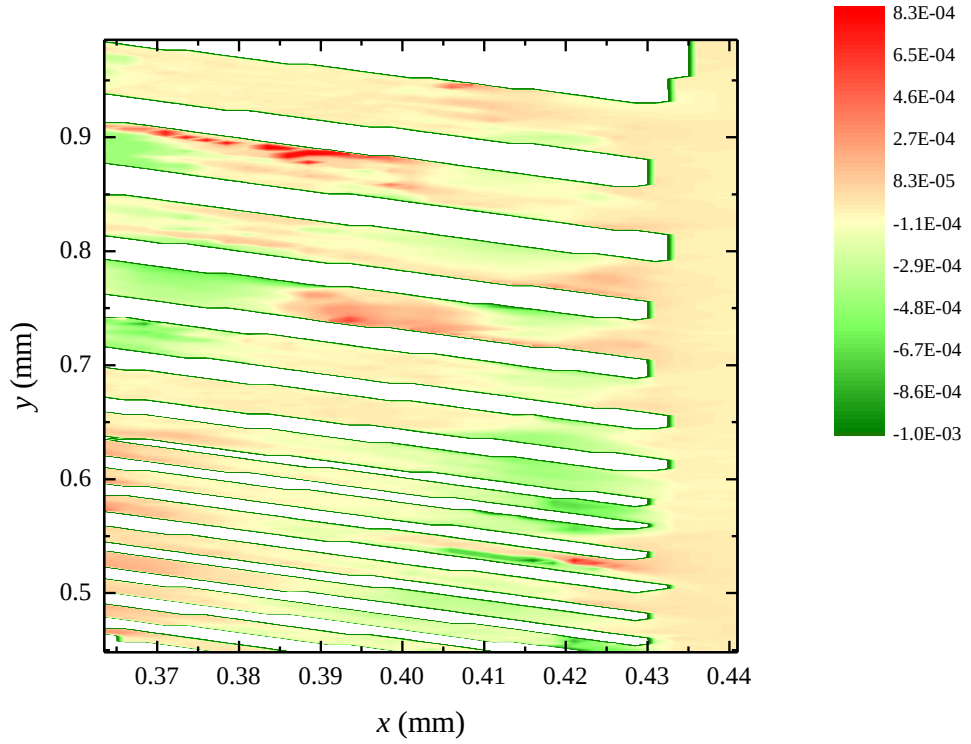


Figure 6.5 Schematic of set of SiB wires showing the level of tilt across them. Areas of increased tilt are shown in red and green.

When comparing solely to the bulk material, adjacent to the wires, it can be seen that there is minimal tilt across them, with a large majority of each wire displaying the same amount of tilt as the bulk. Also, importantly there is no tilt on the bulk SiB. Due to the resolution limit, which is essentially defined by the x-ray beam spot size, the smaller wires Figure are not completely reliable because the signal may be reflected from more than one wire. Due to the divergence of the beam the sample is not always in the focal point of the beam and is therefore not at its smallest size. The result of this is a slightly larger spot size which, for some wires, will encompass more than the thickness of a wire. Therefore a convolution of a couple of the smallest wires may actually be present in the map.

To investigate the reason behind the very small amount of tilt RSMs for three different areas have been explored. To obtain a reference the bulk material RSM is used, as this would serve as an area with superior quality. A second RSM from on the wires was used, from an area which containing the same tilt as the bulk area. The final RSM was from an area where there was a slightly increased degree of tilt.

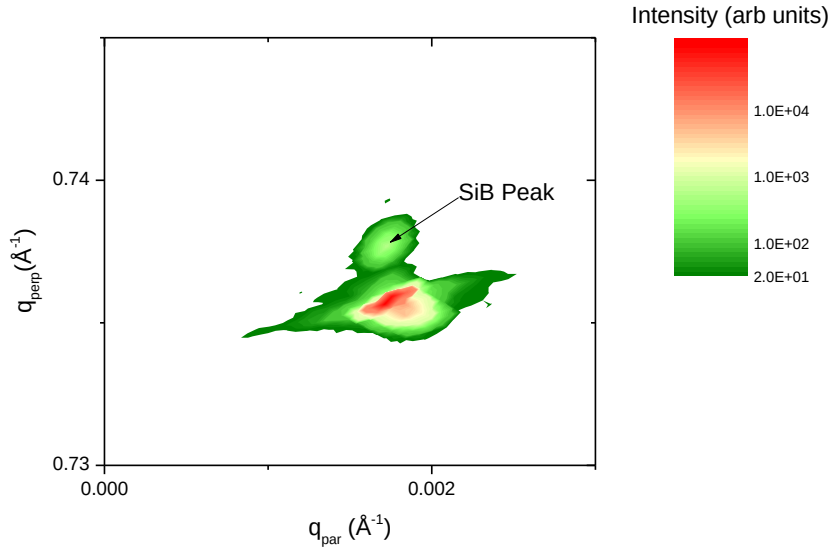


Figure 6.6: RSM taken from the bulk area adjacent to the wires.

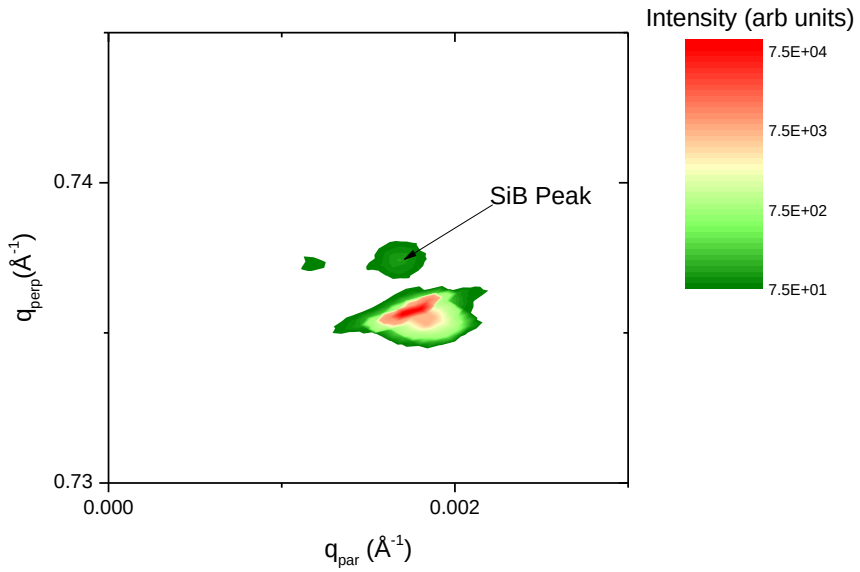


Figure 6.7: RSM taken from the wires where the tilt appears to be the same as the bulk area.

The strongest SiB peak signal is seen in Figure 6.6, which corresponds to the bulk material, which would be expected. The RSMs taken from the wires are displayed in Figures 6.7 and 6.8.

The map which represents the area of increased tilt shows a smearing of the SiB peak when compared to that of the area with no tilt. This smearing of the peak

could cause issues when fitting to the data, and therefore lead to a potentially phantom shift in the tilt. This smearing could be caused by a particulate on the surface, which reduces the intensity of what is already a relatively weak peak in comparison to the Si peak. Figure 6.9 shows evidence of particulates on the surface. This is most likely due to the fact that the conditions for producing the suspended wires were not clean or in a vacuum environment, so therefore is inevitable. However the minimal amount of tilt recorded is not significant enough for the wires to be crooked or warped. As the wires should be composed entirely of SiB the Si peak should be only from the material below the wires, which is why the peak appears essentially the same. It is proposed that while minimal, these small fluctuations could be the result of particulates, and therefore the wires are completely flat.

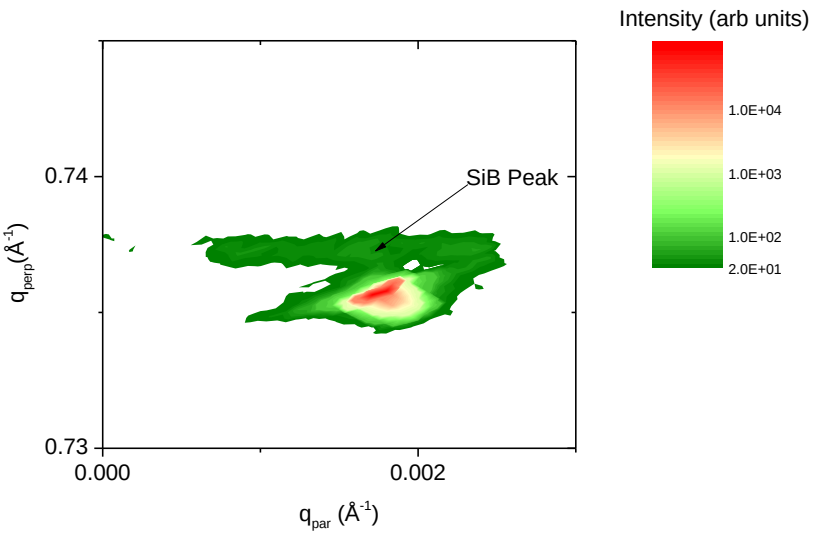


Figure 6.8: RSM taken from the wires where the tilt appears to be different to the bulk area.

(004) measurements were used to investigate the out-of-plane lattice parameter for the wires and the bulk material. It can be seen from Figure 6.10 that there is a slight increase in lattice parameter in the wires. Between the wires and bulk the lattice parameters are 5.419Å and 5.413 Å respectively, which is only a difference of 0.006 Å. The experimental error arises prom the pixel size, and is approximately 0.0001 Å, which is significantly smaller than the measured difference, leading to the conclusion that the difference is measureable. When looking at the

majority areas of the wires which are mostly cream coloured, where the lattice parameter is only around 5.416 Å the difference is even smaller (only three one thousandths of an angstrom). This demonstrates the capability of near Silicon wires completely free of defects, which would otherwise be present if the Silicon was grown on a different material. This therefore has an advantage over using materials such as Germanium, as well as having the doping properties beneficial for other device applications, which pure Silicon does not.

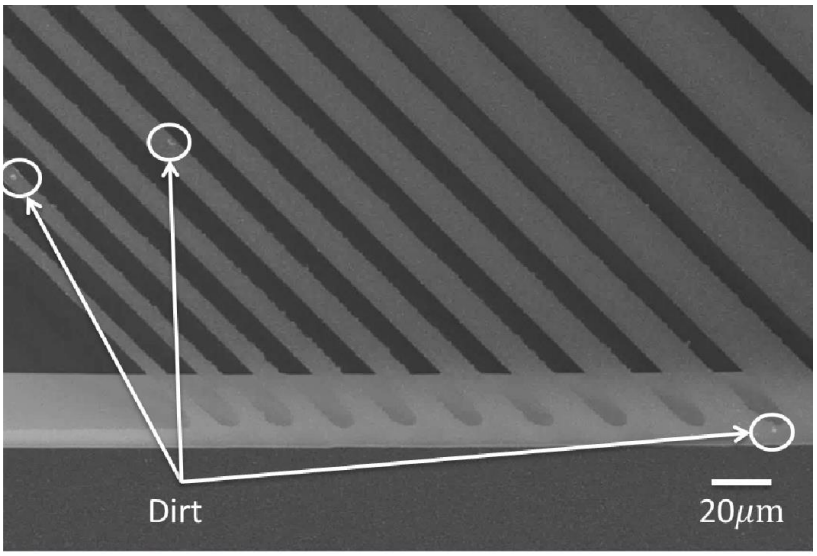


Figure 6.9: SEM image demonstrating an example of some dirt, highlighted by the arrows, on the wires. The dirt is seen by the small lightly coloured regions, which are thought to disrupt the x-ray signal and lead to a potentially phantom tilt.

The bulk areas can be seen on the right edge of the map, with a minimum lattice parameter of 5.413 Å. There are slightly darker areas which correspond to the pseudomorphic region of the bulk. In between these areas, adjacent to the wires the lattice parameter does not appear to be the same as the bulk, and there seems to be a slight relaxation into the bulk. The values of lattice parameter here are between 5.415 Å and 5.414 Å. Although measureable the change in lattice parameter is only 0.002 Å, an order on magnitude larger than the error, which is less than 0.04%. This relaxation looks to be the case only for the larger wires, which would suggest that uniaxially they are acting in a similar way to a membrane by "pulling" some of the bulk material. Next to the smaller wires, circled in blue, it can be noticed that the lattice parameter is more similar to that seen across the larger wires (it is

difficult to comment on specific small wires due to the resolution of the beam). Upon closer inspection of the SEM image of the wires a shelf could be identified.

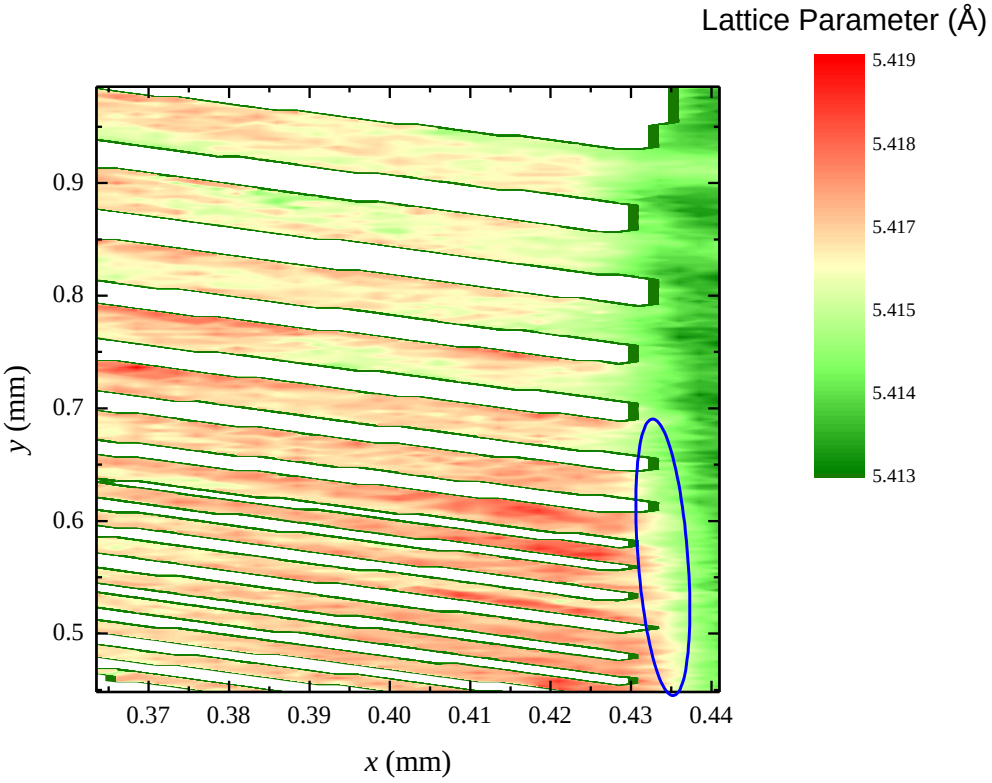


Figure 6.10: Out-of-plane lattice parameter plot from across the wires, using (004) reflection. The general trend is an increase in out-of-plane lattice parameter on the wires which is thought to be caused by their uniaxial nature. Some of the wires do not line up due to the meshing of the data during analysis.

This shelf can be seen in Figure 6.11, where there is a small area of suspended material prior to the bulk material. Therefore it proposed that this area circled in Figure 6.10 is not the bulk material, but actually also suspended. It should be noted that the amount of shelf present is larger for the smaller wires with a reduced shelf seen around the larger wires. Therefore there is a much less noticeable region next to the larger wires with a lattice parameter different to that of the bulk.

Overall the increase in lattice parameter across the wires is very small, which also corresponds to a very small amount of strain relaxation. The smaller wires do show a slight increase in this statistic when compared to the thicker wires, but due to the resolution imposed by the x-ray beam a firm conclusion cannot be made. However,

according to [110] the narrower wires should show more uniaxial strain. More accurately, the width to length ratio of the wires is the important quantity, where a smaller width with equal length (to that of the other wires) should result in a greater amount of strain. Even though the resolution is not excellent for the smaller wires, they do appear to be more heavily strained. The strain across the wires is tensile with respect to the Si substrate

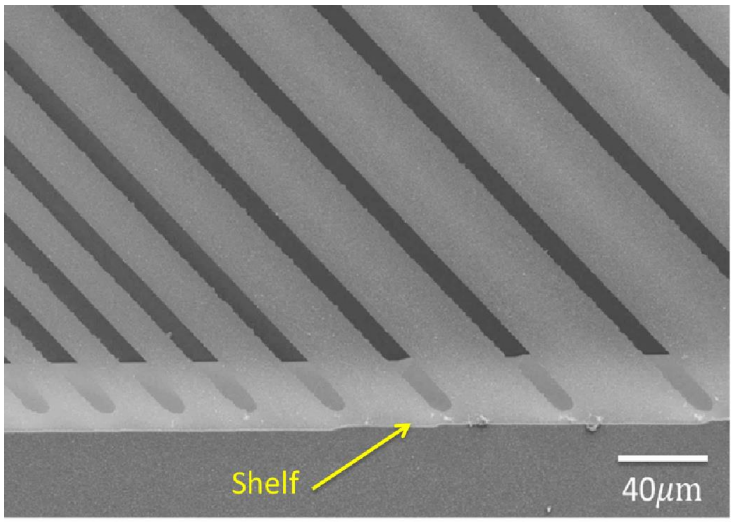


Figure 6.11: SEM image of a set of SiB wires highlighting the suspended shelf adjacent to the wires.

To summarise, it is seen that over the bulk material adjacent to the wires the material is tensile strained, with an out-of-plane lattice parameter less than that of relaxed SiB. It is assumed therefore that the in-plane lattice parameter is that of Silicon. The stretching of this parameter compresses the out-of-plane component. On the suspended wires the out-of-plane lattice parameter increases, and generally maintains this behaviour. This is true for the direction in which the wires were scanned. It is not necessarily true that the behaviour is the same in the other direction, as the strain is presumed to be uniaxial. Although not measured it can be speculated that the in-plane lattice parameter may be expected to decrease as the Silicon substrate is no longer constraining the SiB.

6.4 Summary of SiB wires and membrane

When comparing the two different suspended structures a key point to consider is the type of strain which is imposed on each. The membrane is supported at both

edges and is therefore biaxially strained, whereas the wires are only supported at two points, so are uniaxially strained.

6.4.1 Corner characteristics of membrane

For the biaxially strained membrane it is assumed that the in-plane strain is the same in both the x and y component (and behaves elastically), which means that the effect on the out-of-plane component behaves as expected. Therefore as the x and y in-plane strain pulls the lattice parameter to larger values, the in-plane lattice parameter should decrease. This means that the sample becomes more tensile strained in plane and therefore compressed out of plane. Looking at the trend seen away from the membrane edge and the line scan across the centre of the membrane, it is reasonable to predict that this is the case, and indicative of a SiB membrane. The effect of an increased lattice parameter seen at the edge of the membrane is thought to be caused by a tilting or buckling effect, which is also seen for the Germanium membrane [82]. To investigate this a plot displaying Q_x from the (004) reflection is shown in Figure 6.12.

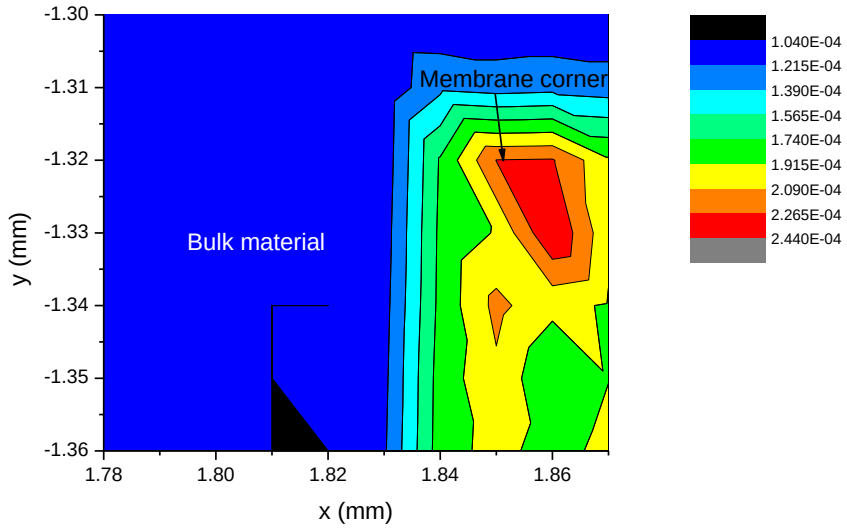


Figure 6.12: (004) map describing the shift in q_x for the corner scan of the membrane. The most amount of tilt is present at the corner of the membrane.

In the (004) reflection the out-of-plane component should change, but the in-plane component should not. It can be seen that there is some change in Q_x on the membrane with the most significant portion around the corner. As the tilt can be

calculated through $tan^{-1}\left(\frac{Q_x}{Q_z}\right)$, it can be seen that a non-zero value of in-plane component corresponds to a tilting of the structure. In this case it is thought that this is caused by a buckling of the membrane. The line scan does show a reduction of lattice parameter more centrally, while the optical microscope image in Figure 6.2 shows a flat membrane. Even though the synchrotron measurements are far better at resolving any tilted features the optical image is useful for observing the middle of the membrane. Also it could be speculated from Figure 6.12 that the tilt may be reducing away from the edge, which is consistent with the behaviour seen for the lattice parameter and Germanium membrane [82].

6.4.2 Comparison of suspended structures

There are two distinct behaviours observed for the lattice parameter variation for each suspended structure analysed in this work. The trend seen for the membrane is an increase in lattice parameter at the edge, before a slight reduction when moving towards the centre. However, for the wires, the lattice parameter increase on to the suspended regions seems to be maintained. An area of interest for the membrane is how the lattice parameter trend continues further from the edge. If the trend observed in Figure 6.4 continues further reduction in the out of plane lattice parameter would be expected.

To confirm the possibility of the SiB membrane acting in a similar way to the Germanium membrane, with regards to lattice parameter, a line scan was obtained which measured across the bulk and suspended layer further into the middle than the map. Figure 6.13 shows the change in lattice parameter over three different areas of the membrane. The blue circle represents the bulk material; the red shows the edge of the membrane, green shows the centre of the membrane. The corner map, shown previously in Figure 6.4, depicts the area of the blue and red rings, which shows the increase in lattice parameter as described.

The proposed decrease can be seen in the line scan, suggesting a similar behaviour to the Germanium counterpart. A fictitious fully relaxed out-of-plane lattice parameter is situated above the top of the line scan. The edge area shows an increased level of strain relaxation, whereas the centre region shows the out-of-plane lattice parameter becoming more strained. Along with the map describing the

shift in q_x it is postulated that the tilt around the edge is a buckling effect, which has led to a degree of strain relaxation. Although time constraints prevented a full strain map including the central region, it seems conclusive from the evidence shown that the membrane shows a very similar profile to that of the Germanium equivalent. There is a small increase in out-of-plane lattice parameter at the frame of the membrane, but overall the lattice parameter across the membrane is smaller than the bulk material.

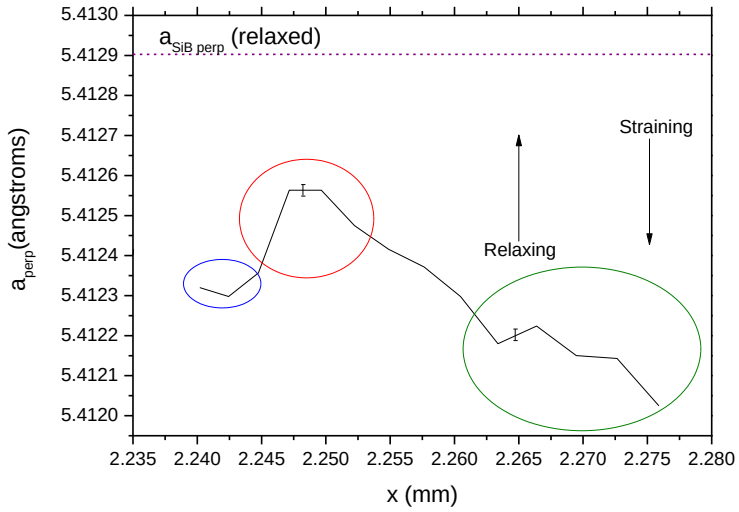


Figure 6.13: Line scan showing lattice parameter variation from edge to centre of membrane. Approximate error bars have been added to show the small deviation across the measurements.

The wires which are uniaxially strained cannot be assumed to act elastically. As the x (y) component is stretched it is not necessarily true that the y (x) component will contract by the same amount. The wires are stretched along the direction of the wires (110 direction). The detachment of the x and y components may suggest that the wires are compressed along the perpendicular direction. This reduction in this direction is compensated by an increase in out of plane lattice parameter, which is manifested in an increase in tensile strain. It could also be observed that the smaller wires exhibit more uniaxial strain than the larger wires, but as previously mentioned the resolution of the technique prevents a firm conclusion. However, as described in [111], it is proposed that as the wire becomes thinner, there is a greater amount of strain, which is consistent.

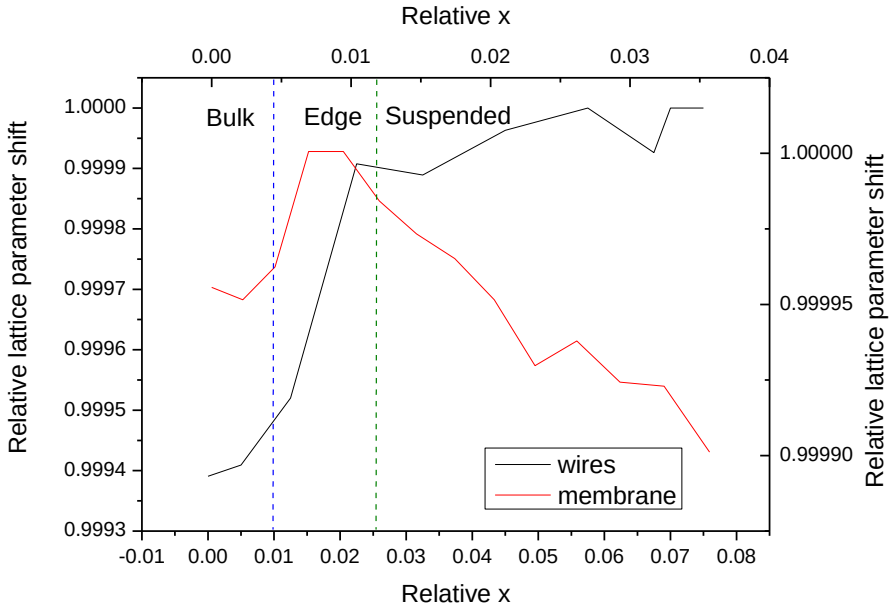


Figure 6.14: Line scans from both wires and membrane. The top and right axis are for the membrane, and the bottom and left axis are the wires. The overall trend for lattice parameter is to increase on the wires and decrease across the membrane.

As can be seen in Figure 6.14 the wires show a different behaviour to the membrane. The variation in lattice parameter in the wires is characterized by an increase followed by a plateau. The plot shows the approximate locations of the bulk material, edge of the suspended material, and the fully suspended material. It should be noted that the lattice parameter shift is relative, and the important information to take is the general shape of each curve. Also the distances are not representative of their actual locations on their respective maps, but instead are a guide for the positions with respect to the suspended material.

Whereas the membrane shows a decrease in lattice parameter when moving into the centre, the wires maintain their increased value. It is thought that there could be biaxial strain at the edge of the wires, which produces the same characteristic as the membrane, however when away from the frame of the structure the strain is uniaxial.

It is important to realize the strain variation, and therefore change in out-of-plane lattice parameter, across both suspended structures is small, which means they could be used for further epitaxial growth. The suspension provides a fast and

inexpensive possibility for a suspended Silicon (99%) layer with minimal strain. As the layer will be grown on Silicon initially, instead of a mismatched material, the layer to be suspended will be perfectly crystalline and defect free. With respect to a SPAD device, there is the potential to grow a structure with a boron doped bottom contact layer, and then remove the substrate. The bottom contact is resistant to the TMAH etch process, which would be very similar to that described for the membrane, and would allow for the etching of the Silicon below. The structure for the suspended SPAD is the reverse (in terms of doping) to the SPAD described in this work. This is to incorporate the etch resistant nature of the p-type dopant to the layer proceeding the substrate. If an n-type dopant such as phosphorus displayed such properties it would be possible to fabricate a suspended device using the original SPAD design.

6.5 Applications and future research

Potential further work with regard to these suspended structures would include testing the etch resistivity of Si:P, fabrication of the SPAD device on such a suspended structure, the application of the reflective coating layer to improve photon detection efficiency. Care would have to be taken when etching the structure, as the nominal doping concentration is slightly lower than that described in this chapter. A possible solution could be to use a highly doped layer followed by a thin Silicon layer before the SPAD structure, or potentially a grading of the doped layer, although neither method has been investigated at this point. As mentioned in chapter 4 a reflective coating could be evaporated onto the underside of the suspended structure as shown in Figure 6.15. As for application to the suspended wire, further optimization would be required. It could be possible to produce an array of SPAD devices along a wire, however fabrication on top of a wire could prove difficult due to its extra fragility.

This method will allow photons which are not absorbed by the Germanium (or the Silicon), to be reflected at the bottom of the device and potentially be absorbed on a second pass through the device.

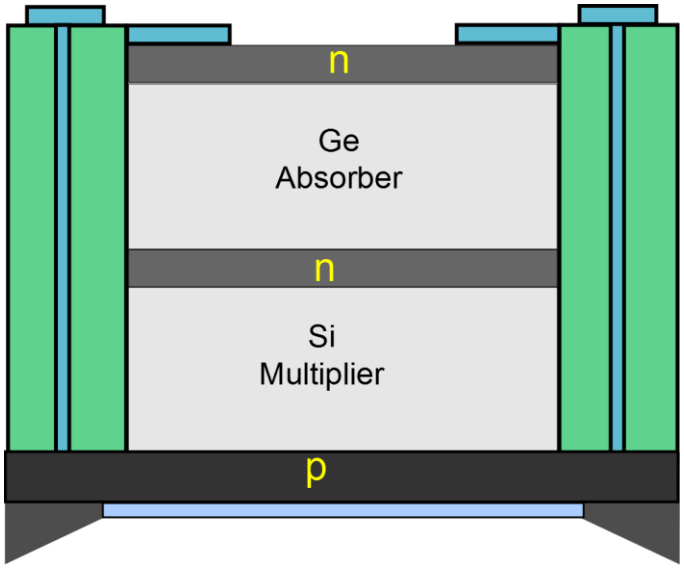


Figure 6.15: Schematic showing the concept of a suspended SPAD device with reflective underside.

Optimization for the gating mode may be required with such a device as there could be a larger error on the arrival time of the photon. The lack of substrate could also provide an improvement to the unwanted parasitic effects. The advantage to using the CVD for such growth, aside from the industrial application, is the superior quality which has been demonstrated in this work at high doping concentrations. For samples grown using MBE, it has been seen that growth quality at high concentrations is an issue and lowering the doping may disrupt the etch resistance required.

7. Thesis Summary

Single Photon Avalanche Diodes operating in the near infrared region have generally been dominated by the III-V compounds when it comes to deciding the optimal material. The only alternative to the III-Vs has been Silicon and Germanium, which since its introduction to this field has been considerably inferior. Early work, particularly using Silicon, was limited by the relatively short wavelength capability imposed by its band gap, while Germanium, whose band gap offered more potential, was severely hampered by the quality of epitaxial growth. With the improvement in epitaxial growth techniques between these two mismatched materials there has become a viable alternative for SPADs. One major advantage of using the Group IV material is the compatibility with other Silicon based technology.

This work introduces the optimization of growth of such a device utilising a separate absorption and multiplication region using Germanium and Silicon respectively. High and low temperature growth of Germanium has led to suitably low threading dislocation densities, which has often hampered device performance. Layers grown during this work have exhibited perfect crystallinity, with relatively low TDD, particularly when compared to other group IV SPAD structures. This has been essential for allowing efficient transport of carriers from absorber region to multiplication region. Observable lower dark counts were measured and attributed to an improvement in growth quality (outlaid in this thesis) and fabrication processing. The specific electric profile through the device has been obtained by accurately controlling doping regions between the layers. While the migration of these atoms can be extremely difficult to completely control, it is critical in ensuring appropriate electric field levels between the two main regions of the device, and therefore allowing for the most effective operation. The growth has been performed using an RP-CVD, which is generally considered the ultimate industry tool for semiconductor growth. While other growth systems could be capable of growing such structures the RP-CVD allows for a fast turnaround. This work has allowed for the realisation and improvement of group IV SPAD devices moving towards an industry scale.

Structures designed during this thesis have been shown to be the first of its kind to detect single photons at a wavelength of 1550 nm. Also huge steps forward have been taken in obtaining comparable detection efficiencies at 1330 nm to the InGaAs/InP

devices. At this wavelength a world leading SPDE of 4% has been measured for a Germanium on Silicon SPAD. While there is room still before these devices may surpass those III-V SPADs, there are some aspects which already make them a more attractive option. Due to the discontinuity between the InGaAs and InP band gaps afterpulsing imposes a limitation on the repetition rate of the device. Devices born through Germanium and Silicon do not suffer as much from this phenomenon, and therefore can be used at a faster rate.

Further work has been performed on the doping of Silicon with boron. Among other things these highly doped layers could be used for low resistance contacts for a SPAD device. The doping concentration achieved in this thesis ($4.5x10^{20}cm^{-3}$) has been the highest reported using an RP-CVD, which again shows their potential in an industry setting. When compared to other growth techniques, dopant concentrations have only exceeded this concentration by compromising the layer quality. The layers grown for this thesis showed perfect crystallinity as well as complete electrical activation, which demonstrates the substitutional nature of the doping above the solubility limit.

A novel and interesting concept has also been introduced by exploring the possibility of suspending a SiB layer. Due to the high level of boron doping it was found that the layer could be resistant to an etching process. This allowed for the removal of the substrate from selected areas and therefore a SiB layer only supported by the edges of a defined square window. Unique analysis of these structures was performed at the Diamond Light Source to obtain structural data using the (004) reflection. Tilt measurements were useful to ensure the layer was still flat and could be used for further growth, or did not suffer from a bowing effect. The lattice parameter of these structures was observed to essentially remain the same as the bulk material strained to the Silicon substrate. This is useful in confirming that the substrate was not necessary. By growing a SPAD structure *ninip* instead of the proposed *pipin* the bottom contact would be etch resistant and could be a candidate for substrate removal. This is a potential avenue for continuation of work in this thesis. In relation to a SPAD the main advantages are two-fold: firstly by removing the substrate any unwanted parasitic conduction could be eliminated, and secondly the introduction of a reflective coat could be employed. This would involve evaporating a reflective coat on to the newly exposed SiB bottom layer. In doing so any photons which were not absorbed while passing through the device could be reflected, subsequently increasing the

detection efficiency of the device. There would be some further optimization required for the growth of the *ninip* structure, using the heavily doped boron bottom contact, as well as the reflective coating, and how that could best be utilised.

There are several avenues to further this work. Further optimization of the SPAD structure could be explored by reducing the segregation further and attempting to reduce the defect density, and therefore DCR. An important aspect of the device to improve is the device fabrication, which has been observed in this work to be a major contributor of dark currents. The concept of suspending the structure could also be explored to improve the SPDE of the SPAD. With regards to the high boron doping concentration and the suspended structures it would be interesting to explore their potential for other device applications as a flat platform for further growth. While the wires are presented as a proof of concept in this work, routes for possible applications could be investigated.

8. References

[1] D. J. Paul, Semiconductor Science and Technology **19**(10) (2004) p. R75

[2] Information from www.semiconductors.org

[3] F. Schaffler **12**(12) (1997) p. 1515

[4] J. Liu et al. **15**(18) (2007) p. 11272

[5] R. R. King et al., Applied Physics Letters, **90**(18) (2007) p. 183516.

[6] J. Michel, J. Liu, and L.C. Kimerling **4**(8) (2010) p. 527

[7] N. Franco, et al., Materials Science and Engineering B-Solid State Materials for Advanced Technology, **124** (2005) p. 123

[8] Kasper et al. Thin Solid Films **520** (2012) p3195

[9] Bellis S, Jackson C and Konig, Laser and Photonic, **5** (2005) p 34

[10] Ashcroft/Mermin, Solid State Physics, (1976)

[11] http://www.sensorsinc.com/technology/what-is-ingaas

[12] Shriake etal. Silicon–Germanium (SiGe) Nanostructures (2011) p29

[13] Lubsandorzhiev, *On the history of photomultiplier tube invention.*

[14] Pavesi, Silicon Photonics Volume 1, (2004)

[15] De Salvador et al., Physical Review B, 2000. **61**(19) p13005

[16] E.Kasper, Properties of strained and relaxed Silicon Germanium (1995) p89

[17] D. Hull and D. J. Bacon, *Introduction to dislocations*. Oxford: Butterworth-Heinemann, 2011.

[18] J. W. Matthews and A. E. Blakeslee, J. Cryst. Growth, **27**, (1974) p 118.

[19] J. C. Bean, J. Vac. Sci. Technol. Vac. Surf. Films **2**(2), (1984) p 436.

[20] Osten et al. Appl Phys Lett **60**(20) (1992)

[21] Shah et al. Solid-State Electronics **62**(2011) p189.

[22] L. H. Hall, J. Electrochem. Soc **119**(11), (1972) p 1593.

[23] D. W. Greve, Mater. Sci. Eng. B, **18**(1), (1993) p. 22–51.

[24] F. C. Frank and J.H.v.d. Mathematical and Physical Sciences, **198**(1053) (1949) p. 205-216.

[25] M. Volmer and A. Weber, Z. Phys. Chem, **119** (1926), p277.

[26] I. N. Stranski, and K. L., Math.-Naturwiss, Kl., Abt. 2B **146** (1938).

[27] Mehrer, Diffusion in Solids, (2007) Pg 413

[28] Frank et al. Diffusion in Crystalline Solids, New York, Academic Press, (1984) p63–142

[29] Greve et al. Journal of electronic materials **21**(6) (1992).

[30] Herzog et al. J. Electrochem. Soc. **131(**12), (1984) p2969

[31] Glass et al. Phys Rev B, **61**(11) (2001)

[32] B. S. Meyerson, Semiconductor Silicon, edited by G. Harbeke and M. J. Schulz, Springer Series in Materials Science **13** (1988), p. 24.

[33] C. P. Parry, et al. J. Appl. Phys. **71(**118) (1992).

[34] S. M. Sze, Physics of Semiconductor Devices, 2nd ed. Wiley, 7643 New York, (1981)

[35] F. N. Schwettman, J. Appl. Phys. **45** (1974) p**.** 1918

[36] G. L. Vick and K. M. Whittle, J. Electrochem. Soc. **119** (1969) p. 1142

[37] Wegrzecka et al. Opto-electronics review **12**(1), (2004) p95.

[38] Colace et al., IEEE Photonics Journal, **1**(2), (2009).

[39] B.K.Lubsandorzhiev, *On the history of photomultiplier tube invention.*

[40] Gower J *Optical Communications Systems* 2nd edn (1993)

[41] 1990 Hamamatsu Photonics, Photomultiplier tubes catalogue

[42] B. T. Levine, C. G. Bethea, and J. C. Campbell, Appl. Phys. Lett. **46**, (1985) p333.

[43] G.S.Buller et al. Meas. Sci. Technol. **21** (2010)

[44] Ringel et al. Materials research society symposium **835** (2005) p211.

[45] Carroll et al. Appl. Phys. Lett. **93**, (2008) p. 183511

[46] Tosi et al. Advanced Photon Counting Techniques II, **6771**, (2007).

[47] G. Vincent, A. Chantre and D J . Bois Appl. Phys. **50 (**1979) p5484.

[48] L. Duraffourg et al. IEEE Journal of quantum electronics 37(1) (2001).

[49] Giovane at al. Applied Physics Letters, **78** (2001) p541.

[50] Itzler et al. Journal of Modern Optics **54** (2007) p283.

[51] W. Maes, et al. Solid State Electron. **33**(6) (1990) p705.

[52] Grant, W. N., *Solid State Electron.* **16**(10) (1973) p1189.

[53] Maes et al. Solis state electronics, **33**(6), (1990) pg705.

[54] Chou et al. Ultramicroscopy **94** (2003) p33.

[55] Hiskett et al. Applied Optics **39**(36) (2000).

[56] S. Cova, M. Ghioni, A. Lacaita, C. Samori and F. Zappa, Appl. Opt. **35 (**1996) p1956.

[57] R.P. Webb, R.J. McIntyre, and J. Conradi RCA Rev. **35**, (1974) p234.

[58] S. Luryi, T. Pearall, H. Tempkin and J. Bean IEEE Electron Device Lett. **7 (**1986) p104.

[59] Louden et al. Optics Lett.Vol. **24**(4) (2002) p219.

[60] Huang et al. *IEEE* Journal of quantum electronics, **43**(3), (2007).

[61] Pelligrini et al. IEEE Journal of Quantum Electronics, **42**(4), (2006) p397.

[62] Liu et al. IEEE, **13**(4) (2007).

[63] Papet et al. Solar Energy Materials & Solar Cells **90** (2006) p2319.

[64] D. Nam et al. Opt Express **19**(27) (2011) p25866.

[65] Rhead et al. Applied Physics Letters, **104**(17) (2014) p172107.

[66] Microsensors, MEMS and Smart Devices (New York: Wiley) p.132

[67] Gardner et al. Microsensors, MEMS and Smart Devices, (2001) p125.

[68] Gardner et al Microsensors, MEMS and Smart Devices (2001) p.

[69] M. Elwenspoek et al. Silicon micromachining. Cambridge: Cambridge University Press (1998).

[70] Li Bet al. Microelectromech Syst, **8**(4) (1999) p366.

[71] Gardner et al Microsensors, MEMS and Smart Devices (2001) p.

[72] V. A. Shahet al. presented at the Ultimate Integration on Silicon (ULIS), (2013) 14th International Conference

[73] Geiger et al., ISTDM 2014 conference preceedings (2014).

[74] Fischetti MV, Laux SE. J Appl Phys. **80** (1996) p2234.

[75] Shah et al. Solid-State Electronics **98** (2014) p93.

[76] Roberts et al. **5**, (2006) p.388.

[77] Digital Instuments, Veeco Training Manual. (2000).

[78] D. Bowen, X-ray metrology in semiconductor manufacturing. Boca Raton CRC/Taylor & Francis, (2006).

[79] D. Bowen, High resolution X-ray diffractometry and topography. Taylor & Francis, (1998).

[80] J. Stangl et al. ChemPhysChem, **10**(17) (2009) p2923.

[81] K. J. S. Sawhney et al. A Test Beamline on Diamond Light Source, (2010) p387.

[82] S. D. Rhead, et al. Appl. Phys. Lett **104**(17) (2014) p172107.

[83] Standard Test Methods for Measuring Resistivity and Hall Coefficient and determining Hall Mobility in Single-Crystal Semiconductors, ASTM

[84] P. Zalm et al. Surface and Interface Analysis, **17**(8) (1991) p556.

[85] MG. Dowsett, 13th International Conference on Secondary Ion Mass Spectrometry and Related Topics (SIMS XIII) (2001).
[86] Borot et al. IEEE, ISTDM (2006).
[87] Shirake, Silicon–Germanium (SiGe) Nanostructures, (2011) p562.
[88] Warburton et al. IEEE, **60**(11), (2013) p3807.
[89] Colace et al. IEEE Photonics Journal, **1**(2) (2009).
[90] Lee et al. Optics express **23**(10) (2015).
[91] Niclass et al. IEEE Journal of selected topics in quantum **13**(4), (2007).
[92] Richardson et al. IEEE Photonics technology letters, **21**(14) (2009).
[93] Lacaita et al. Applied optics **35**(16) (1996).
[94] Liu et al. IEEE Journal of selected topics in quantum electronics **13**(4), (2007).
[95] May et al. Mater. Res. Soc. Symp. Proc. **1039** (2008).
[96] P. Daiet al. Phys. Rev. Lett. **66**, (1991) p1914.
[97] Zappa et al. Sensors and Actuators A **140** (2007) p103.
[98] Bustarret et al. Nature, **444** (2006) p465.
[99] Kirihata etal, Surf. Interface Anal, **40**, (2008) p984.
[100] T. Yamazaki et al. J. Electrochem. Soc. **137**, (1990) p313.
[101] Chen et al. Applied Physics Letters **64**, (1994) p1853.
[102] T. Tatsumi et al. Appl. Phys Lett, **50**, (1987) p1234.
[103] Chen et al. Appl. Phys. Lett. **64**, (1994) p1853.
[104] Weast, CRC Handbook of Chemistry and Physics (1964)
[105] Kucytowski et al. Cryst. Res. Technol. **40**, (2005) p424.
[106] S. Kishino, Adv. X-ray Anal. **16**, (1973) p367.
[107] J. Baribeau and S. Rolfe, Appl. Phys. Lett. **58**, (1991) p2129.
[108] www.microchemicals.com/micro/az_5214e.pdf
[109] O. Tabata et al. Sens. Actuators Phys, **34**(1) (1992) p51.
[110] Minamisawa et al. Nature comms. **3** (2012) p1096.
[111] Geiger et al. Invited Paper ISTDM 2014.